TEKNILLINEN KORKEAKOULU
HELSINKI UNIVERSITY OF TECHNOLOGY
Electroniikan, Tietoliikenteen ja Automaation Tiedekunta
Faculty of Electronics, Communications and Automation

Matthew J. Turnquist

Sub-threshold Operation of a Timing Error Detection Latch

Master's thesis submitted for examination in Espoo 25.05.2009.

Supervisor Professor Kari Halonen

Instructor D.Sc. Lauri Koskinen

TEKNILLINEN KORKEAKOULU

DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Matthew J. Turnquist
Työn nimi:	Sub-threshold Operation of a Timing Error Detection Latch
Päivämäärä:	25.05.2009 **Sivumäärä:** 65
Osasto:	Electroniikan, Tietoliikenteen ja Automaation Tiedekunta
Professuuri:	S-87 Piiritekniikka
Työn valvoja:	Professori Kari Halonen
Työn ohjaaja:	D.Sc. Lauri Koskinen

Ajoitusvirheentunnistus (TED) mahdollistaa energian kulutuksen vähentämisen mikroprosessoreissa. Tässä diplomityössä on kaksi versiota ajoitusvirheentunnistavasta salvasta (esim. TDTBsubI ja TDTBsubII) ja systeemitason testipiiri (SystemTest), joka käyttää TDTBsub salpaa, mikä on suunniteltu toimimaan kynnysalueen alapuolella. Diplomityö esittelee ensin dynaamisen jännitteen skaalauksen (DVS), koska TED käytetään sellaisissa järjestelmissä. Seuraavaksi esitellään teoriaa kynnysalueen alapuolen suunnittelun haasteista. Sitten esitellään molempien TDTBsub salpojen ja SystemTest-lohkojen suunnittelu. Simulaatiotuloksia esitellään keskittyen operaatiotaajuuteen, energian kulutukseen ja toimintavarmuuteen variaatiot huomioon ottaen. Operoitaessa kynnysalueen alapuolella TDTB-piirillä keskityttiin koon mitoittamiseen ja suunnittelutyyliin. Ennen kaikkea kaikkien komponenttien mitoituksen piti olla suurempi kuin minimi CMOS-tekniikan leveydet. Vaikka mitoittamisella saavutettiin toimintavarmuutta kynnysalueen alapuolella toimittaessa myös energian kulutus kasvoi siellä toimittaessa. Perinteisiä vuotovirtojen vähentäviä mitoitustoimenpiteitä tehtiin suurimmalle osalle komponenteista. Logiikkatyyli on tärkeää kynnysalueen alapuolella operoitaessa. TDTBsubII salvassa uuden tekniikan näytetään antavan systeemitason suorituskykyä. Simulaatioilla näytettiin kuinka ajoitusvirheentunnistus kykeni toimimaan kynnystason alapuolella.

TDTBsubI:n ja yhteenlaskun testipiirin piirinkuvio tehtiin 65nm CMOS-prosessilla. TDTBsubII salpaa ei tehty, koska se suunniteltiin piirin määräajan jälkeen. Piiriä tarkasteltaessa osoittautui, että piiri ei toiminut. Piirin toimimattomuus johtui tuotantovaiheessa tapahtuneesta virheestä eikä suunnittelusta.

Avainsanat:	sub-threshold,weak inversion,low power,low voltage,digital CMOS

HELSINKI UNIVERSITY OF TECHNOLOGY

ABSTRACT OF THE MASTER'S THESIS

Author:	Matthew J. Turnquist
Name of the thesis:	Sub-threshold Operation of a Timing Error Detection Latch
Date:	25.05.2009 **Number of pages:** 65
Department:	Faculty of Electronics, Communications and Automation
Professorship:	S-87 Electronic Circuit Design
Supervisor:	Professor Kari Halonen
Instructor:	D.Sc. Lauri Koskinen

Timing error detection (TED) is used to enable the reduction the energy consumption of microprocessors. In this thesis work, two versions of TED latches (i.e. *TDTBsubI* and *TDTBsubII*) and a system-level test circuit (*SystemTest*) that utilizes the *TDTBsubI* latch have been designed to operate in sub-threshold. The thesis first introduces dynamic voltage scaling (DVS) since TED is utilized with such a system. Next, theory is given to highlight the challenges within sub-threshold. The design of the both *TDTBsub* latches and *SystemTest* are then given. Simulation results follow with a focus on operation frequency, energy consumption, and robustness in the presence of variations. To operate *TDTBsub* into sub-threshold, attention was given to sizing and logic style. In general, the sizing of all components was required to be larger than the minimum CMOS width. Although this provided robustness in sub-threshold, the energy consumption in above sub-threshold was much higher. General leakage reduction sizing techniques were also applied to the majority of components. The choice of logic style is important for sub-threshold operation. In the *TDTBsubII* latch, a new technique is shown to provide system-level capability. Simulations displayed the capability of TED in sub-threshold.

The layout of *TDTBsubI* and an adder test circuit were constructed in 65 nm CMOS. The *TDTBsubII* latch was not built since it was designed after the chip deadline. Upon inspection of the chip, it was determined to be inoperative. This mistake was a result of the manufacturering process and not the design in this work.

Keywords: sub-threshold, weak inversion, low power, low voltage, digital CMOS

Foreword

The work for this thesis was carried out at the Electronic Circuit Design Laboratory (ECDL) of TKK. The financial support of this project was provided by the Academy of Finland. The project was coordinated with University of Turku and VTT.

Thanks to all who have helped me along the way. First, I need to thank Professor Kari Halonen for helping me become involved with ECDL. Thanks to my supervisor D.Sc. Lauri Koskinen for his guidance during my work. Thanks to others in the lab who helped me along the way too. Erkka and Jani deserve my gratitude for many late nights and even mornings helping me. Kiitos paljon hyvistä neuvoista Markoselle, Helenalle, Anjalle, ja Anitalle. Thanks to Esa for assisting me in solving all the world's problems. And yes Mika, I do plan to say something today. Kuba ... I need to add a few more marks to the beer counter. Kiitos Karille tiivistelmän takia.

My thanks also goes to my family for supporting my education and goals the entire way. Special thanks to my soon-to-be wife Tiia. Tara was also an inspiration during her days with me at the lab.

Espoossa 25.05.2009

Matthew J. Turnquist

Symbols-and Abbreviations

α	workload
δt	delay of one full-adder
C_g	output capacitance of a characteristic inverter
C_{OX}	gate capacitance per unit area
E_{L1op}	total average energy per operation of LATCH1
E_{LEAKop}	total average leakage energy per operation
E_{LEAK}	leakage energy
E_{SW}	switching energy
E_{TOTop}	total average energy per operation of TDTBsub
E_{TOT}	total energy
I_F	forward source current in EKV model
I_R	reverse drain current in EKV model
I_{dsub1}	first-order approximation of sub-threshold current
I_{dsub2}	EKV approximation of sub-threshold current
I_{OFF}	sub-threshold leakage current between drain and source
I_O	the drain current when V_{GS}=V_T
K_C	threshold voltage coefficient
L_{DP}	critical path depth of an inverter
L_{eff}	effective gate length
S_t	sub-threshold slope
T_{Amax}	the maximum path delay of a full-adder

T_{Amin}	the minimum path delay of a full-adder
T_{CLK}	clock period
t_{dc}	delay from CLK to CLKd
T_{dmax}	maximum delay for logic through a TED system
T_{dmeet}	the delay range of a TED system that guarentees error free operation
T_{dmin}	minimum delay for logic through a TED system
t_{dsub}	propagation delay in sub-threshold
t_d	propagation delay above sub-threshold
T_{op}	time to complete an operation
$t_{r,f}$	rise and fall time above sub-threshold
$t_{subr,f}$	rise or fall time in sub-threshold
T_{valid}	valid region of TED operation
U_T	thermal voltage
V_D	drain voltage of CMOS
V_G	gate voltage of CMOS
V_P	pinch-off voltage of CMOS
V_S	source voltage of CMOS
V_{dd}	supply voltage
V_{TO}	threshold voltage without bias
V_T	threshold voltage
W_{eff}	effective gate width
$W_{n,min}$	minimum NMOS width
$W_{p,min}$	minimum PMOS width
μ_0	zero bias mobility
d	duty cycle
BSIM	Berkeley short-channel IGFET model

DVS	dynamic voltage scaling
EKV	Enz, Krummenacher, and Vittoz
IC	inversion coefficient of EKV model
K	delay fitting parameter
M	mobility temperature exponent
MEP	minimum energy point
N	Number of bits
n	sub-threshold swing coefficient
PoFF	point of first failure
PVT	process,voltage, and temperature conditions
SR	slew rate
TDTB	time-borrowing transition detector
TED	timing error detection
TFD	transient fault detection
VTC	voltage transfer curve

Table of Contents

Chapter 1

Introduction

Significant demand for ultra-low power applications has provided an advantage for circuits capable of sub-threshold operation. The reduction of the supply voltage (V_{dd}) below the threshold voltage (V_T) of transistors, or sub-threshold, provides minimum energy consumption in digital CMOS logic. The ability to scale into sub-threshold is worthwhile for two types of applications. The first is energy constrained systems such as distributed sensor networks [1] and radio frequency identification (RFID) tags. Although the speed of a circuit decreases at low V_{dd}, energy constrained systems do not in fact require high speeds. The second class of applications consists of portable devices that are able to operate at low V_{dd} and slow frequency during times with reduced workload but are still required to switch to larger V_{dd} for high performance. The mobile phone is a primary example of this application since it often enters periods of near idle [2].

In addition to stringent power budgets for energy constrained systems and portable devices, variations must be considered. Increasingly larger variations in process, voltage, and temperature (PVT) conditions causes design uncertainties. In sub-threshold, the effects of PVT are increased substantially. Therefore, circuits capable of low voltage operation need to be adaptable to variations while at the same time be energy optimal. To address both of these needs, dynamic voltage scaling (DVS) has been widely researched and is implemented in a number of processors [3][4][5]. A DVS system operates a circuit at a minimum supply voltage while meeting operation frequency requirements.

To identify and adapt to variations at low V_{dd} and into the sub-threshold region, three adaptive methods utilized within DVS systems have been shown: critical path emulators to track variations [6][7], independent monitoring of variations with individual sensors [8], and scaling supply voltage until failure [9]. The first two methods eliminate a portion of the safety margins due to worst-case PVT conditions but are unable to account for on-chip variation and local changes in temperature and voltage. Unlike the first two methods, the last method accounts for variations at each on-chip location where data is processed.

Timing error detection (TED) is a form of scaling supply voltage until failure and it is used to eliminate all safety margins resulting from worst-case PVT conditions. Using TED,

traditional safety margins are eliminated by lowering V_{dd} up to and even past the point of first-failure (PoFF). At the PoFF, data transition becomes too slow and timing errors occur (e.g. between stages of a pipeline). As the timing error rate increases beyond the PoFF, the recovery energy begins to grow due to the effort required to correct the errors. The tradeoff is between the recovery energy required to fix the errors and quadratic reduction in energy past the traditional safety margin.

Current TED methods, including RazorII [10] and TDTB [11], have shown that adding TED latches to a pipeline can provide significant energy savings. RazorII uses a flip-flop with in situ detection and architectural correction of timing errors. A 64-bit processor in 0.13 μm CMOS was implemented using RazorII and showed a 33% savings in energy. Similar in functionality to RazorII, TDTB was built in 65 nm CMOS and was able to provide 31%-37% energy savings. Although both utilize DVS, neither TED method is capable of operation in the sub-threshold region. However, it is reported in [12] that subthreshold operation would be an ideal application for Razor.

The goal of this thesis is to explore the use of TED in the sub-threshold. More specifically, a version of TDTB (i.e. *TDTBsub*) will be designed to operate in the sub-threshold. Since operation in this region presents inherent design challenges, it will be the focus of this thesis. Chapter 2 first presents the idea of DVS and their utilization of TED. In Chapter 3, the theory behind the digital logic used in constructing *TDTBsub* is given. The majority of the equations from this chapter are referenced numerous times throughout the thesis. Next, in Chapter 4 is presented the design of *TDTBsubI* and *TDTBsubII*, a system-level test circuit, the layout, and a measuring system for the chip. As further explained in Section 4.5, measurement results will not be presented due to a manufacturing process error outside the control of the design in this thesis. To verify the performance of *TDTBsub*, simulations are presented in Chapter 5. A conclusion is given in Chapter 6 which summarizes the thesis work and presents ideas for future work.

Chapter 2

Timing Error Detection (TED)

The objective of this chapter is to describe the operation of TED and give an overview of its history. As mentioned in Chapter 1, TED is an adaptive method used within a DVS system. Therefore, Section 2.1 will provide an overview of DVS systems. The focus of the section is on TED. The fundamentals of TED are then provided in Section 5. Finally, a number of previously implemented TED systems are presented in Section 2.2 to provide a historical perspective of TED. Each of these previous systems uses TED latches within a pipeline structure.

2.1 Dynamic Voltage Scaling (DVS) Overview

For a digital circuit or system to identify and operate at the minimum energy point (MEP), an adaptive system such as DVS is required. An overview of adaptive systems including DVS is given in [13]. DVS has been widely researched and is considered one of the most effective means of reducing energy consumption [3]. A DVS is typically used to provide optimal speed and power performance for microprocessor applications by scaling V_{dd}. This provides significant energy savings due to the quadratic dependence of switching energy with V_{dd} as shown in Eqn. 3.13. An example of a DVS system is shown in Fig. 2.1. The Performance Manager is used to determine the lowest speed needed to meet a desired task (output frequency *Ftar*). The lowest speed, which also takes into account dynamic and process variations (e.g. see Table 3.1), is also used to set the new supply voltage (*Vtar*).

The dynamic and process variations block of Fig. 2.1 provides essential variation data about the system to the Performance Manager. To provide the variation data, three methods utilized *within* DVS systems have been shown:

- DVS_1: critical path emulators to track variations [6][7]
- DVS_2: independent monitoring of variations with individual sensors [8]
- DVS_3: scaling V_{dd} until failure [9][10]

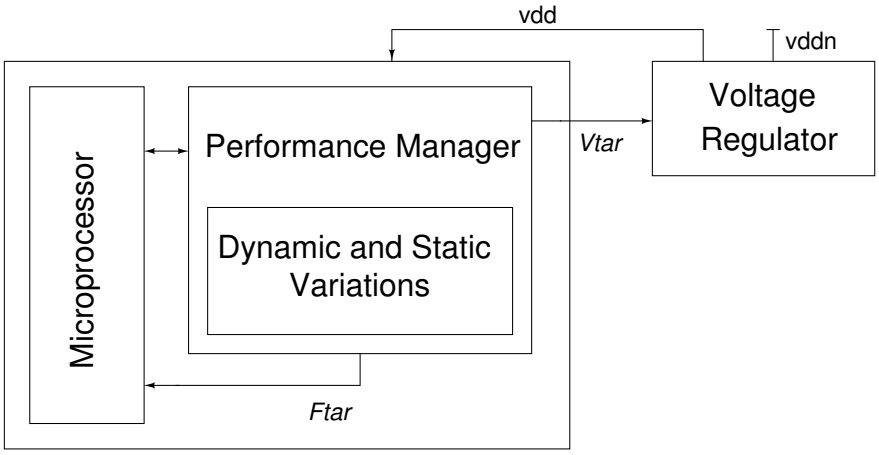


Figure 2.1: High-level view of DVS functionality [4].

The first method, DVS_1, uses critical path emulators to mimic on-chip silicon behavior of the critical path. The emulator is typically a delay-chain that mimics the actual critical path. The delay-chain tracks the critical path delay over local process variations and global fluctuations in V_{dd} and frequency. This eliminates safety margins due to PVT variations. However, the critical path emulator requires safety margins due to other effects. First, the delay-chain does not have the same ambient environment as the critical path since its on-die location is different. Secondly, safety margins are required to address mismatches in scaling characteristics of the critical path [10] and fast-changing transient effects such as coupling noise. In addition to eliminating only some of the safety margins, critical paths are difficult to emulate for complex systems [4]. See [6] for a complex DVS_1 system which is able to operate in sub-threshold.

Independent monitoring with individual sensors, or DVS_2, is another approach to address variations. An example of a DVS_2 system is the TCP core testchip shown in Fig. 2.2. It includes a core, V_{dd} droop sensors, thermal sensors, a dynamic adaptive biasing controller (DAB), distributed noise injectors (i.e. to produce variations), body bias generators, and a PLL unit [8]. To boost the performance of the TCP offload accelerator core, various combinations of V_{dd}, frequency, and body bias are adapted to V_{dd} noise, temperature changes, and transistor aging. The DAB receives the inputs from the sensors and drives the frequency unit, body bias, V_{dd}, and other settings to achieve optimum settings for any given variations. The DAB uses the sensor data as an index to a lookup table preloaded with pre-characterized data representing the optimum settings. Although the TCP core eliminates a portion of the worst-case safety margins, it is unable to account for variation at each of area of the chip in which data is processed.

Unlike DVS_1 and DVS_2, DVS_3 does not have safety margins since it scales V_{dd} until a failure results (Fig. 2.3). At the point of first-failure (PoFF), data transition becomes too slow and timing errors occur. As the timing error rate increases beyond the PoFF, the recovery energy begins to grow due to the effort required to correct the errors. The tradeoff is

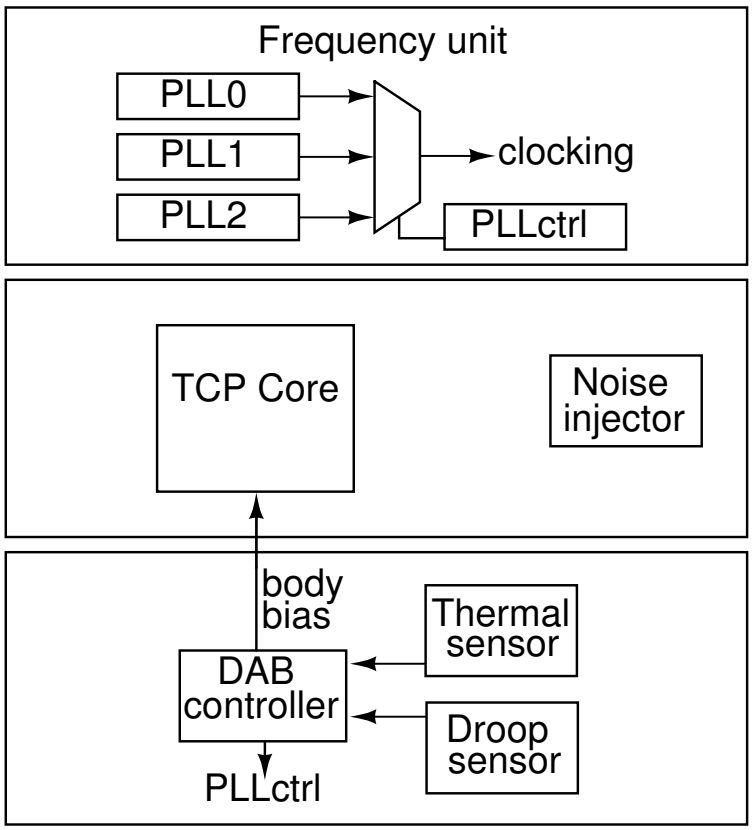


Figure 2.2: TCP core test chip.

between the recovery energy required to fix the errors and quadratic reduction in energy past the traditional safety margin. The fundamental difference between DVS_3 and other systems (i.e. DVS_1 and DVS_2) is that DVS_3 accounts for variations at each on-chip location where data is processed.

2.1.1 TED Fundamentals

An example of DVS_3, called timing error detection (TED), is shown in Fig. 2.4. TED latches are inserted onto all critical paths of a pipeline. The TED latch has the ability to recognize a timing error and generate an error signal when data transition becomes too slow between combinational logic. For example, when the input data *D* to a TED latch transitions under a *CLK* low, the delay due to the value of V_{dd} is considered appropriate and no timing error signals are generated (Fig. 2.5). However, when *D* transitions under the time when *CLK* is high (i.e. the TED window), an *ERROR* is generated. This indicates that the delay due to the value of V_{dd} is too low for the combinational logic and thus timing errors will occur until V_{dd} is increased.

Any *ERROR* signals are passed to an OR gate which forwards any errors to the Voltage Control block. This block determines the acceptable error rate and adjusts V_{dd} within the pipeline accordingly. In the pipeline, V_{dd} is decreased until the PoFF for a given frequency.

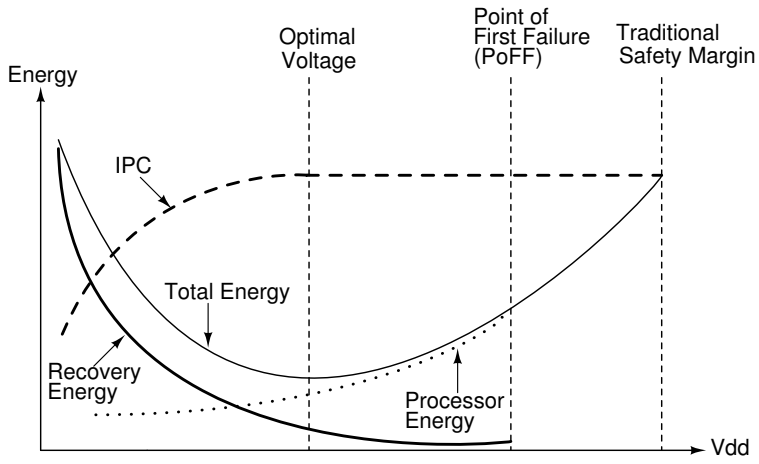


Figure 2.3: Relationships between V_{dd} and energy. As the error rate increases beyond the PoFF, the recovery energy grows due to the effort required in correcting the errors.

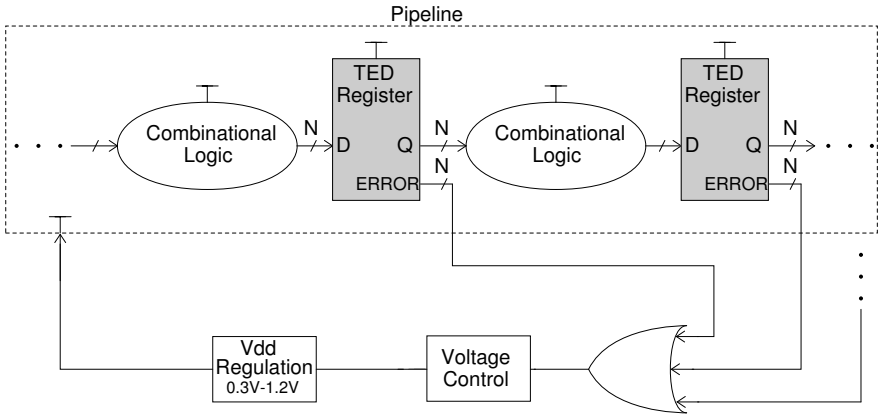


Figure 2.4: System-level view of a generic pipeline system that uses TED latches.

Scaling to the PoFF translates into eliminating all safety margins due to global and local variations thus providing significant energy savings. For TED systems, V_{dd} can be scaled lower than the PoFF, thus providing additional energy savings [10].

2.2 TED Implementations

A number of different approaches to TED have been shown in the past few years. Although all the systems have similar functionality, each has a different approach to the architecture at the circuit and system-level. Three DVS_3 systems will be presented here. The last system presented, TDTB, is the circuit used as the starting point for this thesis so an understanding of its operation is crucial. It should be noted that none of the systems presented here are capable of sub-threshold region operation.

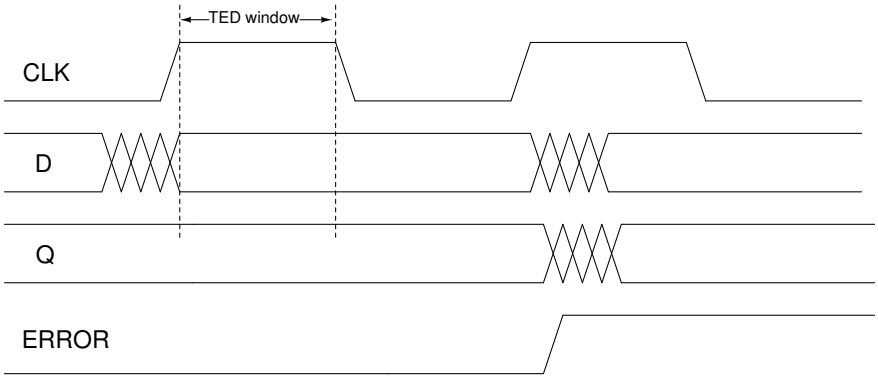


Figure 2.5: TED timing diagram. An *ERROR* signal results when *D* transitions under a *CLK* high.

2.2.1 Original

The TED technique is first introduced in [14] and is called transient fault detection (TFD). TFD is used to detect timing errors known to be the result of local variations and soft errors due to energetic particles. The TFD was constructed using two latches and a comparator (Fig. 2.6). If a data transition (out) occurs more than once during a period of *CLK* (i.e. due to a timing error or soft error), an error signal (err) results. For this case, the data transition is not latched in LATCH. Since the Extra LATCH has a delayed *CLK* signal (i.e. *CLK*+X), it successfully latches the data. As a result, the comparator detects the difference in Q1 and Q2 and generates *err*. To address manage the *err* signals, [14] propose that the most economical solution is to use a TED technique combined with an architectural replay procedure. This means that each time an *err* is generated, the data should be again placed through the combinational logic.

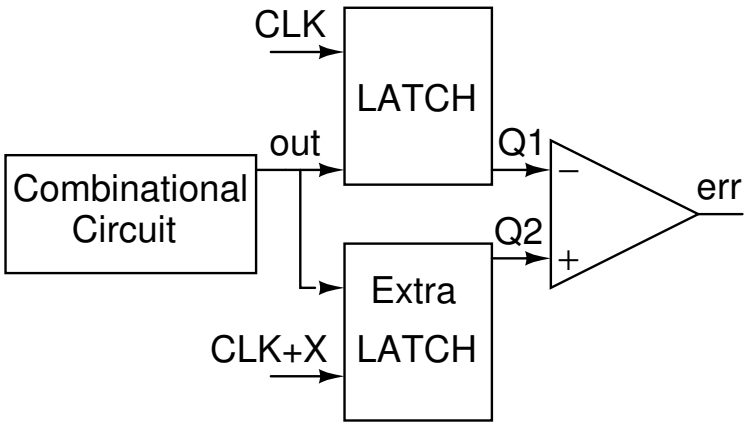


Figure 2.6: First known TED implementation [14].

2.2.2 Razor

Razor II [15] is a DVS system that uses a Razor II flip-flop for TED. The (delay-error tolerant) flip-flops from Razor II are placed on all critical paths within a pipeline and are used to scale the supply voltage V_{dd} to the point of first failure (PoFF). V_{dd} can actually be scaled below the PoFF, deliberately tolerating a target error rate, thus further reducing energy consumption. A timing error is considered a trade-off between the overhead of error-correction (i.e. architectural replay) and the additional energy savings due to operation below PoFF. Although Razor II operates in strong inversion, it was reported in [12] that subthreshold voltage scaling would be an ideal application for Razor.

The architecture of Razor II is shown in Fig. 2.7. It uses a single positive level-sensitive latch, with the addition of a transition-detector (TD) controlled by a detection clock (DC). Data is considered on time when *D* transitions prior to the rising edge of the *CLK*. However, if *D* transitions after the rising edge of *CLK*, during transparency, then the transition of latch node N occurs when TD is enabled and an *ERROR* is generated. At the rising edge, DC provides a short pulse to disable TD for at least a delay of the clock-to-Q delay of the latch. Razor II uses 47 transistors. The power overhead for a Razor II flip-flop compared to standard flip-flop for a 10% activity factor is 28.5%. The overhead is calculated from a 1.2 V, 0.13 μm CMOS circuit [15].

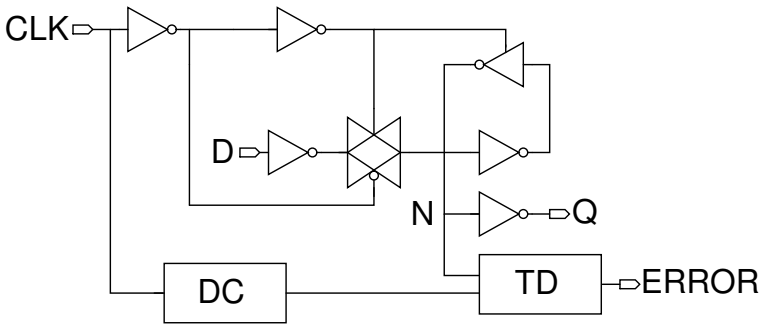


Figure 2.7: Razor II circuit [15].

2.2.3 Time-Borrowing Transition Detector (TDTB)

TDTB is shown in Fig. 2.8. The TDTB monitors the location of input data *D* transitions with respect to the *CLK*. For each *D* transition, a pulse at the output of the XOR results. During the time when *CLK* is low, node K is driven high by P1 thus keeping *ERROR* low. The pulse from the XOR has no effect during this time. A late arriving *D*, during the time when *CLK* is logic high, provides a path for K to be discharged since both N1 and N2 are ON. Consequently, the *ERROR* node transitions to a logic high.

The transition detector may become metastable but is not considered a hindrance. Metastability may occur if *D* arrives close to a *CLK* edge, which is the boundary of a timing failure.

The LATCH is transparent during the *CLK* high and thus *D* still propagates to the next stage. The error buffer will be driven high or low during metastability, therefore, either case still maintains correct functionality [11].

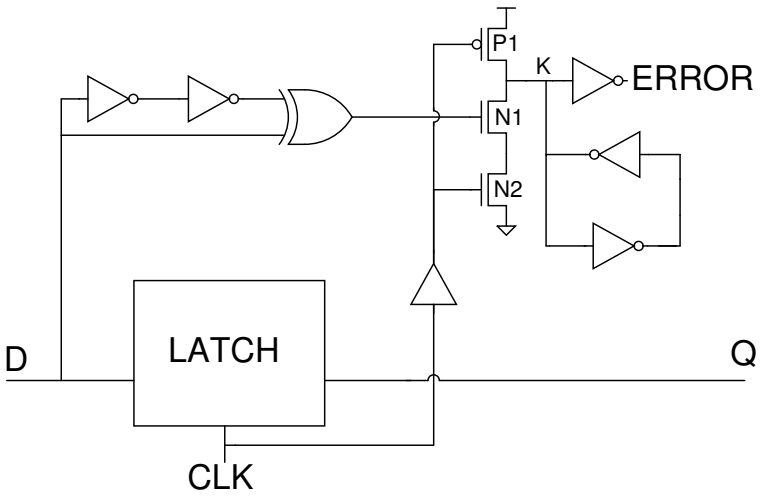


Figure 2.8: TDTB Circuit [11].

The TDTB was compared to two other TEDs, including a version of Razor found in [16], and it was reported that TDTB showed much less energy overhead [11]. The concept of TDTB is extended in this Master's thesis by exploring its use in the sub-threshold region. Our design includes a version of TDTB that operates in sub-threshold and is referred to as *TDTBsub* throughout this paper.

Chapter 3

Theory

This chapter presents the theory behind the digital logic used in constructing *TDTBsub*. There are numerous digital logic styles in existence today and they are well documented in [17]. One of the most ubiquitous styles is called static CMOS. This style ensures that the output of a gate is in steady-state and always a logical function of the inputs independent of time [18]. Static CMOS typically[1] uses a pull-down network (PDN) and/or pull-up network (PUN) to generate the output of a gate. The PDN and PUN are constructed from NMOS and PMOS transistors, respectively. These transistors are the focus of this chapter since the design of *TDTBsub* used only static CMOS.

Section 3.1 presents useful conduction, timing, and leakage equations for *TDTBsub*. A more detailed look at static CMOS with respect to variations is then presented in Section 3.2. Both Section 3.1 and 3.2 provide a background for understanding the minimum energy consumption of CMOS in Section 3.3.

3.1 Static CMOS

Static CMOS transistor equations describing the conduction, timing, and leakage equations are given in this section. These were found to be the most insightful equations for the design of *TDTBsub*. Throughout the remainder of this thesis, static CMOS is referred to as CMOS.

3.1.1 Conduction

To first understand conduction in CMOS, a minimum sized NMOS transistor with a constant drain voltage of 1.2 V was simulated for different values of V_{GS} (Fig. 3.1). Each of the CMOS I_{ds} operation regions are labeled in the figure. The current I_{ds} has linear dependency in the strong inversion region while in the moderate inversion region it shows quadratic dependency. The current in the sub-threshold region does not drop abruptly below V_{GS}=V_T, but actually decays exponentially, similar to BJT operation. The effect of CMOS transistors

[1] PDN and/or PUN is used for complementary and ratioed static CMOS. Pass-transistor logic is also form of static CMOS but does not use a PDN or PUN [17].

conducting below the threshold voltage (V_T) is called sub-threshold conduction. Unlike moderate and strong inversion, in which the drift component of current dominates, sub-threshold conduction is dominated by diffusion current [19]. Sub-threshold conduction can be expressed by two useful equations. First, a simple first-order approximation is shown by [2]:

$$I_{dsub1} = I_O exp(\frac{V_{GS} - V_T}{nU_T}), \tag{3.1}$$

where n is the sub-threshold swing coefficient, the thermal voltage is defined as $U_T = kT/q$ (25 mV for 25 C), and I_O is the drain current when V_{GS}=V_T:

$$I_O = \mu_0 C_{OX} \frac{W_{eff}}{L_{eff}} (n-1) U_T^2, \tag{3.2}$$

where μ_0 is the zero bias mobility, C_{OX} is the gate capacitance per unit area, and W_{eff} and L_{eff} are the effective gate width and length, respectively.

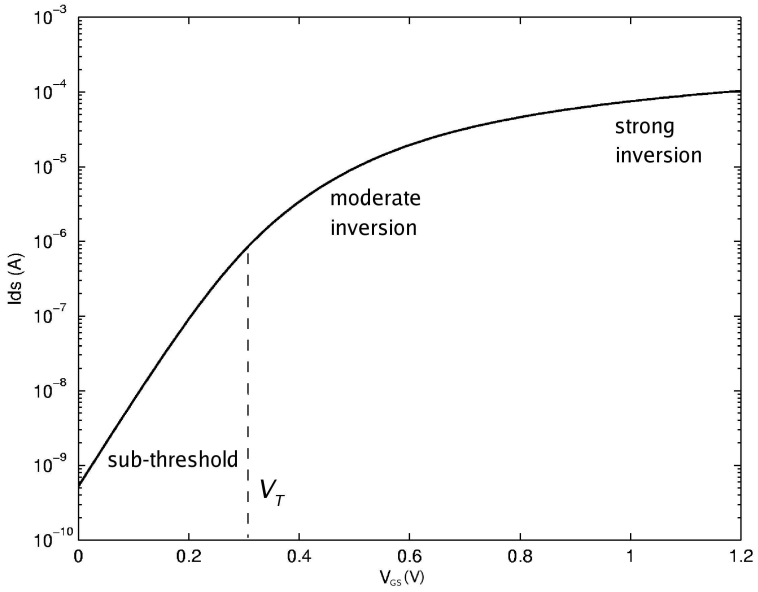


Figure 3.1: CMOS I_{ds} operation regions shown for an NMOS with V_{ds}=1.2 V and V_{GS} swept from 0 V to 1.2 V.

A more detailed and intuitive current equation applicable for the sub-threshold region, moderate inversion, and strong inversion is found from the Enz, Krummenacher, and Vittoz (EKV) model. It provides simple hand calculations and a small amount of parameters for calculation of the current. The model was specially developed for low-voltage and/or low-current circuit design. Its roots derive from the design of analog circuits used within the first electronic watches. The model has been used primarily in the design of low power analog circuits, but it also finds application in digital logic [20]. For a thorough presentation of its influential history see [21].

The main transistor design parameter of the EKV model is called the inversion coefficient (IC). The IC replaces the long-time used overdrive voltage, which works well for the strong

inversion region but not in the sub-threshold region[21]. Since the forward current at the source (I_F) and the reverse current at the drain (I_R) are used within the IC, they must first be defined [2]:

$$I_{F,R} = I_{spec} ln^2(1 + exp(\frac{v_p - v_{s,d}}{2})), \tag{3.3}$$

where I_{spec}=$2n\mu_O C_{OX} W_{eff}/L_{eff} U_T^2$. $I_{F,R}$ is normalized to $i_{f,r}$ by dividing (3.3) by I_{spec}. The terms $v_{s,d}$ and v_p refer to the normalized versions of $V_{S,D}$ and the pinch-off voltage, respectively. The total current in any region of operation can then be generated from (3.3) since $I_{ds} = I_F - I_R$. The total current in the sub-threshold region is therefore:

$$I_{dsub2} = I_{spec} exp(\frac{V_G - V_{TO}}{nU_T})(exp(\frac{-V_S}{U_T}) - exp(\frac{-V_D}{U_T})). \tag{3.4}$$

Equation (3.4) highlights the symmetry of CMOS by showing that the total current is dependent on the superposition of each terminal voltage. For stacks of CMOS devices, (3.4) is particularly useful in providing a first-order hand calculation to understand the operation. It has proved to be useful for operation of digital circuits in sub-threshold [20].

Since the definition of $i_{f,r}$ has already been shown, the IC can be found. As shown in Fig. 3.2, the IC is determined by its location in the (i_f,i_r) plane. If i_f > 1 and i_r > 1, then both components are in the strong inversion region and the whole channel is strongly inverted. When i_f > 1 and i_r < 1, the reverse current at the drain is negligible. If i_f < 1 and i_r > 1, the forward current at the source is negligible. If i_f < 1 and i_r < 1, then both components are in the sub-threshold region. The channel is nearly depleted free of electrons and holes but a small amount of electrons exists in the channel. The moderate inversion region is at the point when $i_{f,r}$=1 ($I_{F,R}$=I_{spec}) [21]. To summarize, a transistor is in the sub-threshold region when IC$\ll$ 1, the moderate inversion region for IC$\cong$ 1, and the strong inversion region for IC$\gg$1.

3.1.2 Timing

As described by the previous conduction equations (3.1),(3.3), and (3.4), the output conduction of CMOS is reduced as V_{dd} decreases. Decreasing conduction has a significant effect on the propagation delay and the rise and fall times $t_{r,f}$. The propagation delay for an inverter operating in strong inversion is [2]:

$$t_d = \frac{KC_g V_{dd}}{(V_{dd} - V_T)^\alpha}, \tag{3.5}$$

where K is a delay-fitting parameter, α is the workload, and C_g is the output capacitance of a characteristic inverter. The denominator of (3.5) models the current of the characteristic inverter above operation in the sub-threshold region and causes t_d to increase linearly with smaller V_{dd}.

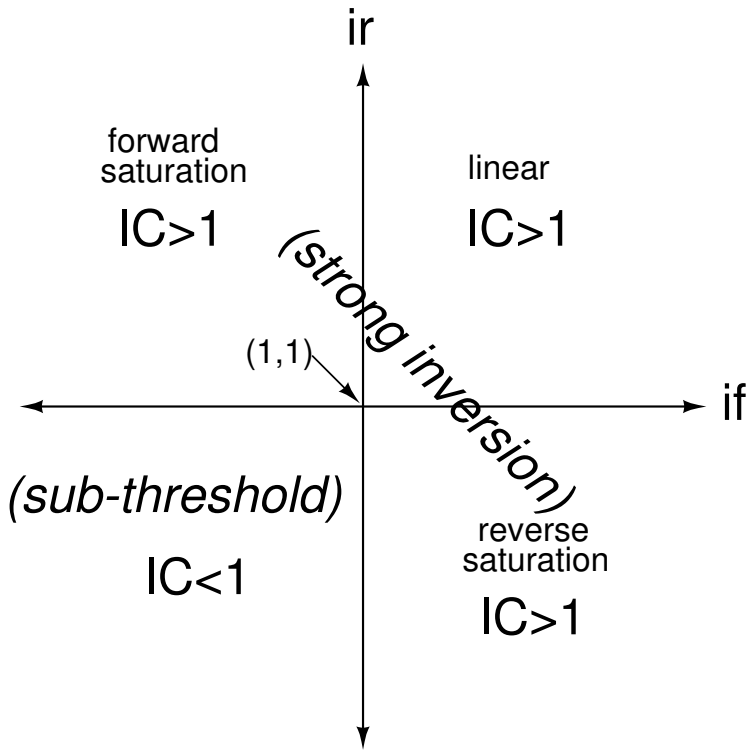


Figure 3.2: MOS transistor operation modes [11].

In subthreshold, the propagation delay of an inverter is [2]:

$$t_{d,sub} = \frac{KC_gV_{dd}}{I_o exp(\frac{V_{dd}-V_T}{nU_T})}. \tag{3.6}$$

Similarly to (3.5), the denominator of (3.6) models the current of the characteristic inverter. In the sub-threshold region, the delay grows exponentially with decreasing V_{dd}.

Derivation of $t_{r,f}$ for an inverter is similar to the approach used for the delay. Here will be presented an approximation to $t_{r,f}$ in the sub-threshold region. The slew rate (SR) or the maximum rate of change of output voltage can be defined as [22]:

$$SR = \frac{C_g}{I_{ds}}. \tag{3.7}$$

Assuming that the $t_{r,f}$ of a signal has a linear slope and is simulated from rail-to-rail, SR can be regarded as approximately equivalent to $t_{r,f}$. Using the sub-threshold current equation from (3.1), $t_{r,f}$ can then be defined for the sub-threshold region as:

$$tsub_{r,f} = \frac{C_g}{I_O exp(\frac{V_{r,f}-V_T}{nU_T})}, \tag{3.8}$$

where $V_{r,f}$ defines the voltage of a PMOS (i.e. V_r=-V_{dd}) and NMOS (i.e. V_f=V_{dd}). As in (3.6), the $t_{r,f}$ grows exponentially as V_{dd} decreases.

3.1.3 Leakage Current Mechanisms

In addition to conduction and timing equations, the leakage should be considered in designing the CMOS logic of *TDTBsub*. Scaling transistors sizes into the nanometer regime has caused dramatic increases in CMOS leakage current (I_{OFF}) due to reductions in V_T, channel length (L), and gate oxide thickness (t_{ox}). Increased I_{OFF} has become a major portion (i.e. 30 - 50%) of the total power consumption in many scaled technologies [23]. Furthermore, as V_{dd} is reduced to sub-threshold region levels, the propagation delay increases exponentially (3.6), which leads to an exponential increase in leakage energy. The leakage power is thus integrated over much longer periods of time and as a result leakage energy begins to dominate switching energy. Although leakage can occur during active and standby modes of CMOS, this thesis addresses only standby mode (I_{OFF}).

The five main short-channel leakage mechanisms are shown in Fig. 3.3. *Ia* is the pn junction tunneling leakage. As previously stated, I_{CH} is the sub-threshold leakage. *Ig* is the gate leakage and it consists of leakage from gate-to-channel, gate-to-drain, and gate-to-source. *Ie* is the gate-induced barrier lowering (GIDL) leakage. The channel punchthrough current is *If*. *Ia* and *Ig* occur when the transistor is ON and OFF. I_{CH}, *Ie*, and *If* are all OFF state currents [19].

The general consensus is that I_{CH} and *Ig* are the most dominant leakage types in nanometer digital CMOS [23]. For a short channel device (i.e. 65 nm), the amount of I_{CH} is a function of the drain voltage [19]. The source and drain depletion width in the vertical direction have a strong effect on the band bending due to the short L. As the drain bias changes, the depletion width also changes therefore causing V_T to vary. As previously explained, a change in V_T results in the variation of sub-threshold leakage current. The effect is called Drain-induced barrier lowering (DIBL). Throughout the work in this thesis, I_{CH} was found to be the most troublesome leakage mechanism in sub-threshold. The sub-threshold leakage can be expressed as [23]:

$$I_{CH} = \mu_0 C_{OX} \frac{W_{eff} U_T^2}{L_{eff}} e^{1.8} e^{[(V_{GS}-V_T)/(nU_T)]} [1 - e^{-V_{ds}/U_t}]. \tag{3.9}$$

Gate leakage *Ig* contributes significantly to I_{OFF} in nanometer digital CMOS. As CMOS lengths scale below 60 nm, it is expected to be the dominant form of leakage [24]. However, gate leakage depends strongly on the operation region. In the sub-threshold and moderate inversion region, gate leakage to the channel (Fig. 3.3) is negligible while leakage from the gate-to-drain and gate-to-source is more dominant. Above sub-threshold, the larger electric field across the oxide allows for carriers to tunnel through the oxide [2].

To further understand leakage in CMOS, a 65 nm NMOS transistor was simulated at a typical process corner. A simulation for a LVT (low threshold voltage), SVT (standard threshold voltage), and HVT (high threshold voltage) NMOS transistor is shown in Fig. 3.4. From Fig. 3.4, a number of important device parameters within sub-threshold can also be

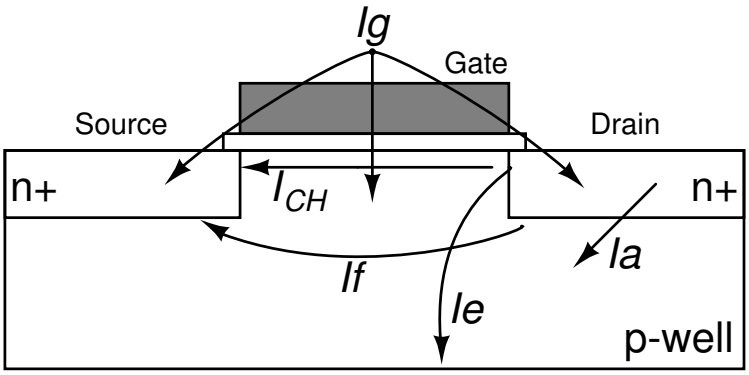


Figure 3.3: The 5 main leakage mechanisms [19].

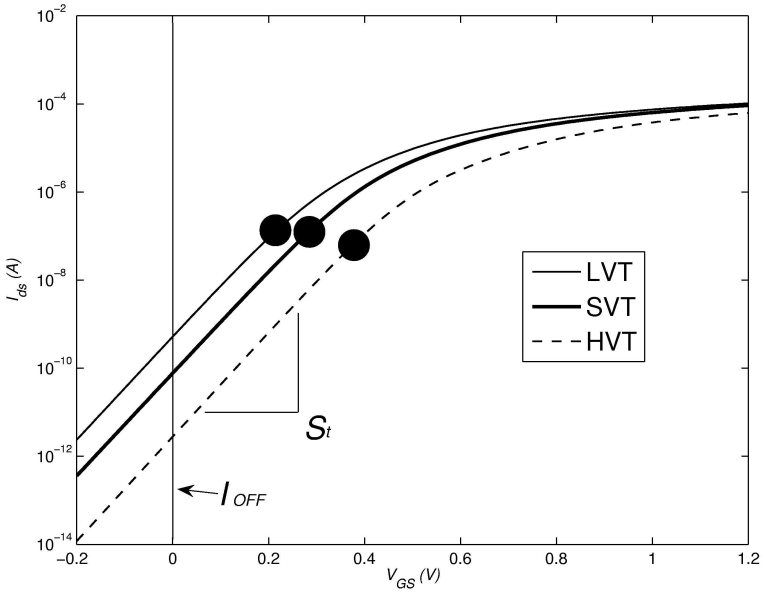


Figure 3.4: Sub-threshold leakage of an LVT, SVT, and HVT 65 nm NMOS transistor. The drain to source was a constant 1.2 V and the source tied to ground. The circles represent the location of V_T on the V_{GS} axis for each I_{ds} curve.

recognized including I_{OFF}, V_T, and sub-threshold slope (S_t). I_{OFF} is the current between the drain and source and consists of leakage from both I_{CH} and Ig when V_{GS} is 0 V. The circles represent the location of V_T on the V_{GS} axis for each threshold type. It is imperative to recognize that as V_T is decreased, the leakage increases. S_t indicates how effectively the NMOS (or PMOS) transistor is turned off. It describes the inverse of the slope of I_{ds} vs V_{GS} and is considered a quality measure of a device. The definition of S_t is:

$$S_t = nU_T ln10, \tag{3.10}$$

where S_t is expressed in mV/decade. An ideal transistor gives n=1 and S_t=60 mV/decade at room temperature. In practice however, n is greater than 1 for actual devices [17]. From Fig. 3.4, the sub-threshold slope was calculated as approximately 85 mV/decade for each I_{ds}.

3.2 Variation in CMOS

Accounting for variation in addition to conduction, timing, and leakage of CMOS presents additional challenges. Modern CMOS nanometer-scale technologies have a large number of process and dynamic variations to consider (Table 3.1). Process variations and their impact will first be discussed followed by the effects of temperature.

Table 3.1: Sources of variability in nanometer-scale technologies [25]

Process Variations	**Dynamic Variations**	**Simulation Tools**
channel length	temperature	timing analysis
channel width	supply voltage	rc extraction
threshold voltage	aging	cell modeling I-V curves
overlap capacitance	cross-coupling capacitance	circuit simulations
nesting effects	multiple input switching	process files
interconnect	workload	transistor models
	average supply voltage	

3.2.1 Process Variation

Careful consideration of sizing and logic style within the sub-threshold region is required to account for process variations at low voltages and deep sub-micron technologies (i.e. CMOS 65 nm). Process variation can be divided into global and local variations. Global variations affect all devices on a wafer similarly (i.e. discrepancies in alignment) with an effect seen in the sub-threshold region as strong PMOS or weak NMOS, or vice verse. Local variations affect devices on the same wafer differently and consist of both systematic and random components. Typically, global variations have been of most concern in digital CMOS design. As the size of transistors length (L) has decreased though, local variations have grown larger than global variations [26].

The local variation of most concern is the mismatch between devices on the same die and is typically modeled as a difference between their threshold voltages (V_T). The standard deviation of V_T variation is approximately proportional to $(WL)^{-1/2}$ [2]. Since device currents have an exponential dependence on V_T (3.1), V_T variation causes changes in delay, energy consumption, and output swings due to changes in I_{ON}/I_{OFF} [2]. Considering the effects of V_T, larger sized devices are required for circuits to operate in the sub-threshold region.

Logic styles are affected differently by variations. For example, ratioed circuits using sizing ratios to guarantee functionality are not robust in the sub-threshold region. Since CMOS device current depends exponentially on V_T in the sub-threshold region, local variation becomes significant relative to the sizes of devices and results in incorrect functionality for ratioed circuits [2]. Circuits with many parallel paths should also be avoided since they are less robust to global variations in sub-threshold.

A useful sizing metric for different logic styles in 65 nm was provided in [2] and is repeated in Table 3.2. It can be used to achieve robustness in the sub-threshold region for global and local variations of a number of different logic styles. For convenience, the values are normalized to the minimum width of the process. The sizing metric was used as a starting point in sizing many of the circuits in this thesis, being generated from Monte-Carlo simulations that were performed to vary V_T at each global and temperature corner. The device widths were increased until 99.9% of samples had proper output swing.

Table 3.2: Variation-driven sizing metric for 65 nm [2]

	Min. size constraint
1 NMOS in pull-down	2.67
2 series NMOS in pull-down	5.33
3 series NMOS in pull-down	6.33
Transmission gate	2
1 PMOS in pull-up	1

3.2.2 Temperature Effects

Temperature has an effect on all CMOS I_{ds} operation regions. In this section, two fundamental temperature equations will first be presented. Next, the temperature effects on the delay and noise margins of an inverter will be given. Examining the threshold voltage and mobility as a function of temperature help to clarify the effect of temperature on CMOS [27]:

$$V_T(T) = V_T(T_0) - K_C T, \tag{3.11}$$

$$\mu(T) = \mu(T_0)(\frac{T}{T_0})^{-M}, \tag{3.12}$$

where T_0=300 K, K_C is the threshold voltage coefficient (typical value 2.4 mV/K) and M is the mobility temperature exponent (typically near 1.5). In strong inversion, lower mobility dominates and results in slower circuits for high temperatures. In the sub-threshold region, a lower V_T dominates [2] for high temperatures and results in a lower delay.

The decrease in delay in the sub-threshold region can be confirmed from the results of rise and fall time simulations for a CMOS inverter (Table 3.3). A voltage from the strong inversion and sub-threshold region were simulated for FF and SS global process corners. An FF corner provides low temperature and high V_{dd}, while an SS corner gives high temperature and low V_{dd}. The temperature is within the exponent of (3.8) and thus considered more dominant than V_{dd}. In strong inversion, an SS corner gives a larger $t_{r,f}$ than FF as is expected from (3.7). Unlike strong inversion, an SS corner in the sub-threshold region gives a smaller $t_{r,f}$ than FF as expected from (3.8). Since V_{dd} is less than V_T, the exponent of (3.8) is negative. Thus, increases in temperature and decreases in V_{dd} for SS both contribute to a

smaller $t_{r,f}$. In summary, the delay effects of SS and FF corners are opposite in the sub-threshold region to those in strong inversion.

Table 3.3: *Rise and fall times of an inverter.*

V_{dd} (V)	t_r (sec.)	t_f (sec.)
1.2 SS	1.68e-11	1.10e-11
1.2 FF	1.17e-11	0.69e-11
0.4 SS	6.67e-8	6.03e-8
0.4 FF	8.85e-8	7.74e-8

Although the delay of an inverter is affected by temperature, the noise margins do not show a significant dependency. The effects of temperature on an inverter operating in all CMOS I_{ds} operation regions has little effect on voltage swing. The effects of temperature on an inverter at V_{dd}=0.3 V was performed in [28]. The nominal output low (V_{OL}) and output high (V_{OH}) voltages of the inverter from 0 C to 100 C degraded only slightly and the overall effect was considered negligible. The standard deviation of V_{OH} and V_{OL} with local V_T variation increased slightly at high temperature and from this it was implied that the high temperature corner was the worst case for sub-threshold region functionality.

3.3 Energy Consumption

Understanding the effects of CMOS in the sub-threshold region in the previous sections is important since the minimum energy per operation point (MEP) of CMOS occurs in the sub-threshold region. The total energy per operation of a digital circuit consists of two components: switching and leakage energy [2]. The switching energy assuming rail-to-rail swing is:

$$E_{SW} = C_{eff} V_{DD}^2, \tag{3.13}$$

where C_{eff} is the average effective switched capacitance per operation.

The leakage energy is:

$$E_{LEAK} = (I_{LEAK} V_{dd}) T_{op} \tag{3.14}$$

$$= W_{eff} K C_g L_{DP} V_{DD}^2 exp(\frac{-V_{dd}}{nU_T}), \tag{3.15}$$

where T_{op} is the time to complete an operation and L_{DP} is the critical path depth in characteristic inverter delays.

The total energy per operation is then expressed as [2]:

$$E_{TOT} = E_{SW} + E_L \tag{3.16}$$

$$= V_{DD}^2 [C_{eff} + W_{eff} K C_g L_{DP} exp(-V_{dd}/nU_T)]. \tag{3.17}$$

The set of equations above can be used to solve the optimum energy consumption and highlight the importance of major parameters with regards to the optimum point. E_{SW} is dominant for most V_{dd} but as the voltage scales to sub-threshold region levels, E_{LEAK} starts to dominate the energy consumption for low V_{dd} thus providing the MEP. The effects of both E_{SW} and E_{LEAK} can be seen from the MEP simulations in Chapter 5. As expected, the MEP occurred in the sub-threshold region for each of these simulations.

To define the V_{dd} at which the MEP should occur, the derivative of (3.17) is taken with respect to V_{dd}, setting it equal to zero, and applying a number of rearrangements the optimal V_{dd} is solved [2]:

$$V_{DDopt} = nV_T(2 - lambertW(\frac{-2C_{eff}}{W_{eff}KC_gL_{DP}})). \tag{3.18}$$

The *lambertW* function is used to solve equations involving exponentials and is described in [29].

Chapter 4

Design

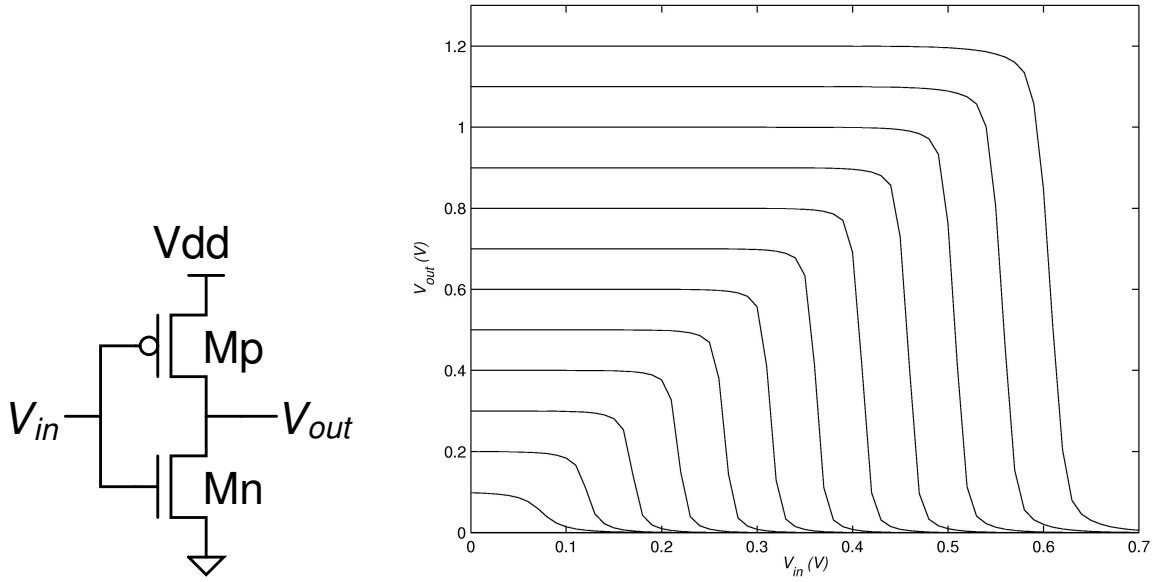
This chapter provides the design of a TED latch capable of operation in sub-threshold, moderate inversion, and strong inversion. The TED latch is called *TDTBsub* and was built using a 65 nm CMOS process. It was simulated using the BSIM4 transistor model and eldo. Since there are common factors in the design of *TDTBsubI* and *TDTBsubII*, *TDTBsub* refers to both versions. Similarly to Chapter 3, the focus of this chapter is on the design in the sub-threshold region due its inherent challenges.

Since the inverter is the most frequently used device in *TDTBsub*, its operation and sizing requirements are given in Section 4.1. Next, the functionality, circuit style, leakage, and system-level considerations of *TDTBsubI* are introduced in Section 4.2. A similar presentation follows for *TDTBsubII* in Section 4.3. *TDTBsubII* not only works in sub-threshold, but it gives a new approach to reset the timing error signals. A system-level test circuit (i.e. *SystemTest1*) is given in Section 4.4 to make use of the *TDTBsubI* latch at the system-level. The pre- and post-manufacturing layout of *SystemTest1* and *TDTBsubI* latch follows in Section 4.5. Finally, the design of the measurement system is provided in Section 4.6.

4.1 Inverter

A push-pull inverter, or inverter, is a key component in the design of *TDTBsub*. The inverter is the most common component in *TDTBsub* and, therefore, its operation has a large impact on overall robustness, energy consumption, and the minimal voltage operation point. An inverter is shown in Fig. 4.1 (a.). The size of Mp's and Mn's transistors width are called W_p and W_n, respectively. W_n was sized to $W_{n,min}$ (0.135 μm) while W_p was 1.5*$W_{n,min}$. The transistors length (L) was increased from the minimum of 60 nm to 80 nm for both Mp and Mn. This is applied to the majority of devices in this chapter since Long-L transistors (i.e. nominal + 10%) have less leakage [30]. As shown in the voltage transfer curve (VTC) of Fig. 4.1 (b.), the inverter maintains correct functionality deep into the sub-threshold region.

The size of K=(W_p/W_n) limits the functionality of CMOS circuits as V_{dd} is lowered. Fig. 4.2 shows the minimum V_{dd} in which an inverter with HVT NMOS and PMOS maintains a



(a) (b)

Figure 4.1: (a.) Inverter schematic (b.) VTC for V_{dd} from 0 to 1.2 V. W_n=0.135μm and W_p=1.5*W_n.

10-90% output swing. The size of W_p was swept while W_n was sized to the $W_{n,min}$. The maximum W_p is a result of leakage through Mp which works to keep V_{out} high as Mn is working to correctly drive it low. The maximum W_p displays the maximum PMOS width to ensure V_{out} is 10% or less of V_{dd}. The minimum W_p curve is due to the leakage through Mn which works to keep V_{out} incorrectly low as V_{in} transitions low. The minimum W_p curves denotes the minimum PMOS width in which V_{out} is 90% or more of V_{dd}. From Fig. 4.2, the minimum operating voltage of 70 mV was found from the intersection of the minimum W_p and maximum W_p. The inverter used to generate Fig. 4.2 had a larger minimum operating voltage compared to the minimum operating voltage of 50 mV from the 18 μm process used in [2] for a similar simulation. This indicates that the leakage is greater for the 65 nm used in this simulation even for an HVT device which is designed to have lower leakage. A larger leakage translates into the minimum W_p curve shifting up and the maximum W_p curve shifting down.

The minimum operating voltage can also be examined in terms of an inverters drive strength. The amount of current delivered by the NMOS and PMOS is called the drive strength. The minimum operating voltage occurs when the NMOS and PMOS transistors of an inverter have the same current. For the drive strength simulations, an inverter was used with NMOS sized as $W_{n,min}$ while PMOS as W_p=K*$W_{n,min}$. To understand the drive strength, a number of different K were used as shown in Table 4.1. The average current, I_{AVG}, was simulated for 10 transitions of V_{in}. Next, the NMOS and PMOS current I_N and I_P, respectively, were simulated. To understand the differences in current between I_N and

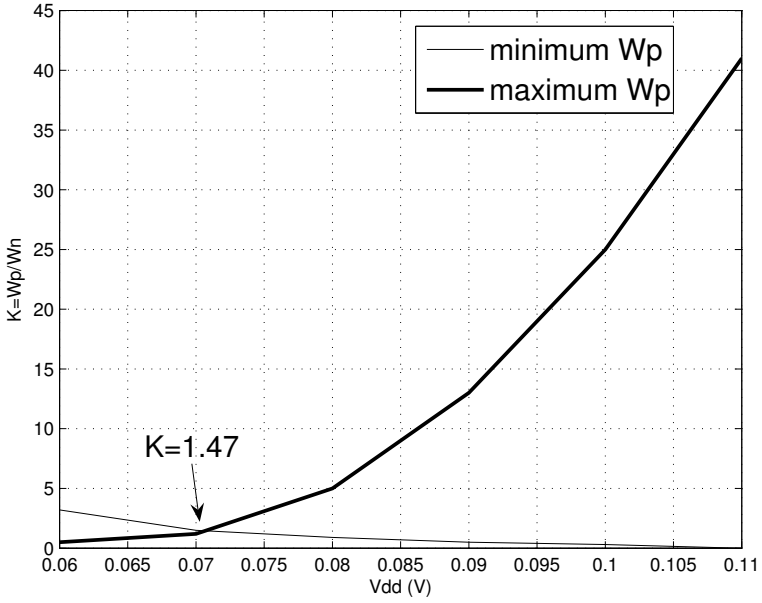


Figure 4.2: Inverter with HVT devices for typical process corners. The minimum operating voltage is 70 mV.

I_P as a percentage of I_{AVG}, the following equation was applied:

$$I_{diff} = \frac{I_N - |I_P|}{I_{AVG}}. \tag{4.1}$$

The absolute value of (4.1) is displayed in Table 4.1 and it shows that the smallest percentage difference in drive strength within the sub-threshold region was at K=1.5. Therefore, using K=1.47 to obtain a minimum operating voltage as already suggested in Fig. 4.2, is confirmed by simulating the drive strength of the inverter.

Table 4.1: $|I_{diff}|$

V_{dd} (V)	**K=1**	**K=1.5**	**2**
1.2	1.28%	5.42%	8.87%
0.6	5.78%	7.29%	8.29%
0.3	1.22%	0.79%	1.06%

In addition to having a low voltage operation of an inverter, the switching threshold of an inverter (V_M), should be located around the middle of the available voltage swing (i.e. V_{dd}/2), since this provides comparable values for the low and high noise margins [17]. The values of V_M were simulated as shown in Table 4.2. At sub-threshold region levels, K=1.5 provides a reasonable value of V_M. Since the data from Table 4.1 and 4.2 provided low voltage operation and good noise margins, the majority of inverters within *TDTBsub* were sized using K=1.5.

Table 4.2: V_M as a function of K and V_{dd}

V_{dd} (V)	**K=1**	**K=1.5**	**2**
1.2	0.577	0.586	0.590
0.6	0.283	0.287	0.288
0.3	0.142	0.145	0.146

4.2 *TDTBsubI*

A detailed explanation describing the functionality of each device within *TDTBsubI* is first presented in Section 4.2.1. For devices to operate correctly in the sub-threshold region, careful attention must be given to the sizing and circuit style as explained in Section 4.2.2. The leakage effects on *TDTBsubI* is then given in Section 4.3.3. Finally, system-level considerations are provided in Section 4.2.4.

4.2.1 Functionality

TDTBsubI, shown in Fig. 4.3, consists of a cell-library positive edge triggered latch (LATCH1) for passing data *D* between combinations logic stages, an additional positive edge triggered latch (LATCH2) for holding an *ERROR* signal high until the next *CLK* rising edge, and a transition detector (TD) for detecting transitions of *D*. The TD consists of four main components: a pulse generator, a pull-down network, a Keeper, and a CLK-delay chain (CDC). The TD ensures an appropriate V_{dd} for a given processing task. For example, if V_{dd} is too low, the data passed between the combinations logic states of Fig. 2.4 in Chapter 3 may become slow and timing errors may occur. At the system-level, a timing error requires an increase in V_{dd}.

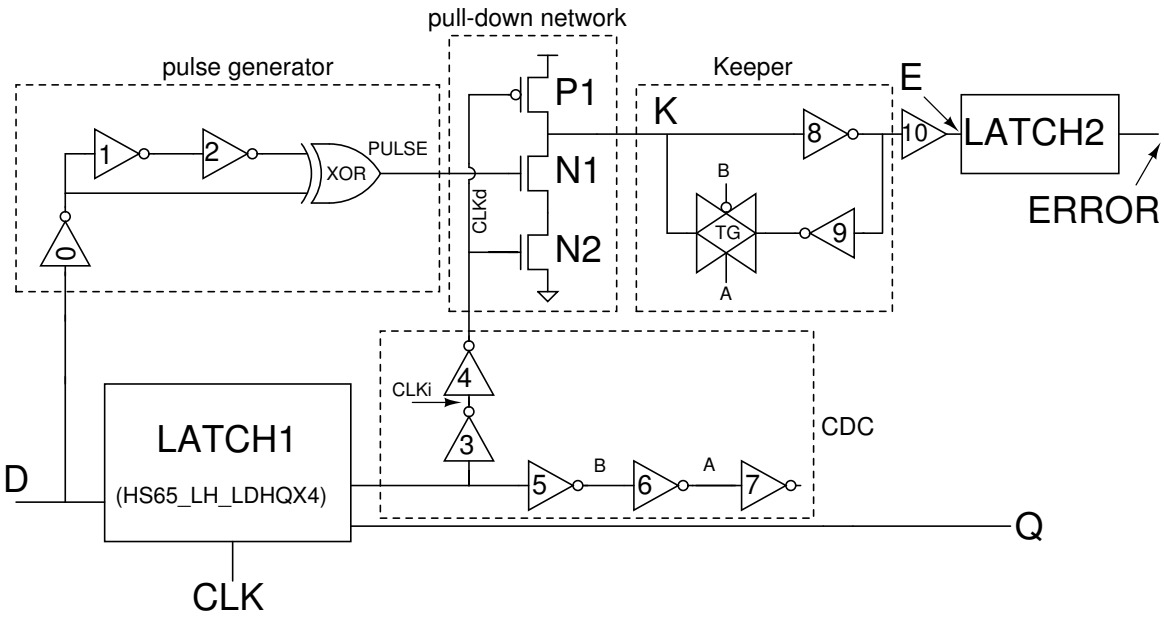


Figure 4.3: *TDTBsubI*.

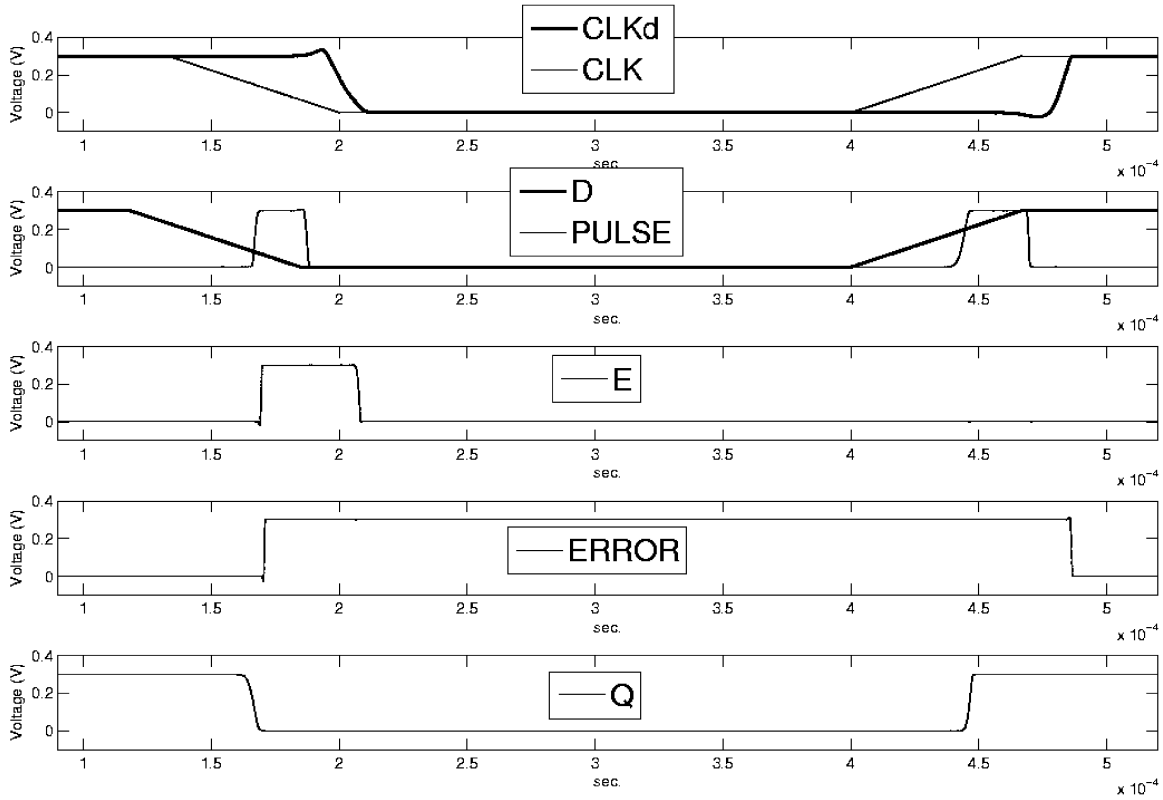


Figure 4.4: *TDTBsubI* operation at 0.3 V for typical process parameters. When *D* transitions under *CLKd* high, which means *D* is arriving too late, an *ERROR* signal results. No *ERROR* signal results when *D* transitions and *CLKd* is low. Note that the *ERROR* signal is reset on the *CLKd* rising edge.

The pulse generator uses inverters 0,1,2 and an XOR to generate a short voltage pulse (*PULSE*) each time the data (*D*) transitions (Fig. 4.4). Inverter 0 is needed to ensure the XOR's pulses have a similar duration and delay (from the *D* transition) as the $t_{r,f}$ time of *D* changes. Inverters 1 and 2 determine the duration of *PULSE*. A *PULSE* with small delay and small duration is desirable to ensure that when *D* transitions closely before the delayed clock signal (*CLKd*) transitions from low to high, an *ERROR* is *not* generated. For example, see the *D* transition located at 0.4 ms in Fig. 4.4. If the *PULSE* duration is too long, it would fall under the *CLKd* high signal and incorrectly generate an *ERROR* even though *D* transitioned at the same time as *CLK*.

The pull-down network is controlled by *CLKd* which allows node K to transition low if *D* transitions at the same time *CLKd* is high. Inverters 3 and 4 determine the delay of *CLKd*. P1 is used to reset node K high and subsequently ensure *ERROR* is low when *CLKd* is low (i.e. *ERROR*-reset). N1 is turned ON when *D* transitions and *CLKd* is high. N2 allows for K to be connected to ground if an XOR pulse occurs when *CLKd* is high.

The Keeper is used to prevent a floating node and it switches states if node K is driven low through the pull-down network. For K to operate correctly in the sub-threshold region, the Keeper utilizes a transmission gate (TG) and is considered a key component of *TDTBsubI*. In the sub-threshold region, the large sensitivity to process variations of a ratioed circuit (i.e. Keeper without a TG) creates operational problems [2]. Inverters 4 and 5 are used to emulate the delay of *CLKi* and *CLKd*. This resulted in less variation of *CLKd* and better operational performance.

The CDC sizing is important to the accuracy of timing error detection. Since the location of *PULSE* varies with respect to the rising/falling edge of *D*, the delay *CLKd* created by inverters 3 and 4 must account for the worst-case *PULSE* delay. The XOR pulse delay changes for each V_{dd}. Thus, *CLKd* must be delayed long enough to account for the longest XOR pulse as V_{dd} changes.

4.2.2 Sizing and Circuit Style

For operation of *TDTBsubI* in sub-threshold, the sizing and circuit style was examined for each element of the TD (i.e. pulse generator, pull-down network, Keeper, and CDC). The size used for each device is shown in Table 4.3. Since the sizing and circuit style have strong interdependence, they will be discussed together.

The pulse generator contains three inverters (0,1,2) and an XOR from the 65 nm cell-library. Inverter 0 was sized large to ensure that an XOR pulse would be generated for different $t_{r,f}$ times of *D*. Its size was over 7 times larger than recommended from the sizing metric shown in Table 3.2. Inverters 1 and 2 had both large W and L to provide little variation in the duration of the XOR pulse when simulated with 1000 point Monte-Carlo simulations. A number of different sub-threshold region optimized XORs were built and tested, but an optimally sized 65 nm cell-library XOR (HS65 LHS XOR2x6) was chosen for the final

design. The size of the XOR cannot be too large since it adds a large parasitic capacitance to the gate of N1 thus presenting a larger time constant to *PULSE*.

Table 4.3: Device sizing in *TDTBsubI*

Device	**PMOS W/L**	**NMOS W/L**
inverter 0	6.5/0.06	2.7/0.06
inverter 1=2	6.5/0.35	2.7/0.35
inverter 3=4=5	6.0/0.5	5.7/0.5
inverter 6=7	5.1/0.08	1.72/0.08
P1	4.0/0.08	N/A
N1	N/A	2.0/0.08
N2	N/A	4.0/0.08
TG	0.405/0.08	0.405/0.08

In the pull-down network, transistors N1 and N2 are used to pull down node K when *D* transitions under a *CLKd* high. As recommended in Table 3.2, two series 65 nm NMOS devices in pull-down should be sized to W=5.33*W_{Nmin} in order to reduce global and local variation. The local V_T variation causes stacks of devices to exhibit higher variability in output levels. Additionally, stack of devices exhibit decreased drive strength. Assuming I_F (3.3) is larger than I_R, then an decrease in drain voltage for the bottom transistor of the stack causes the drive strength (3.4) to decrease. To ensure a consistent and strong enough current in pulling down K, both devices were initially sized to W=5.33*W_{Nmin}. The size of N1 and N2 were further increased to drive K low faster. Making N2 larger than N1 further increases the speed without affecting the capacitance at the *PULSE* node.

In strong inversion, the proper functionality of the Keeper depends on the sizing ratios of inverters 8 and 9. As mentioned earlier (3.1), device currents have an exponential dependence on V_T in the sub-threshold region, and as a result, variations become significant relative to device sizes [2]. This effect in the sub-threshold region was confirmed via a 50-point Monte-Carlo simulation as shown in Fig. 4.5. When *CLK* goes low, P1 is unable to drive K high for the process corner of strong NMOS and weak PMOS. The stronger (NMOS) inverter 9 works against the weaker P1 (PMOS) and keeps K incorrectly low. Making P1 extremely large is an option, but this creates additional problems such as increased leakage and unwanted delay to *CLKd*. The best solution is to close the feedback path to inverter 9 when P1 needs to reset K high by use of a TG. The keepers utilization of the TG is a key component of *TDTBsubI*. Additional Monte-Carlo simulations in Section 5.2.3 confirm that this non-ratioed circuit style presents increased robustness in the sub-threshold region.

At the output of the Keeper, a buffer and LATCH2 were placed to ensure system-level timing requirements were achieved. As shown in Fig. 2.4, the *ERROR* signals of the *TDTB-sub* latches are fed to an OR gate. To ensure that the OR gate catches all timing errors, *E* should be high until the next cycle of *CLK*. This was achieved by adding the buffer and LATCH2. The buffer ensured that *E* was delayed long enough to meet the setup time of

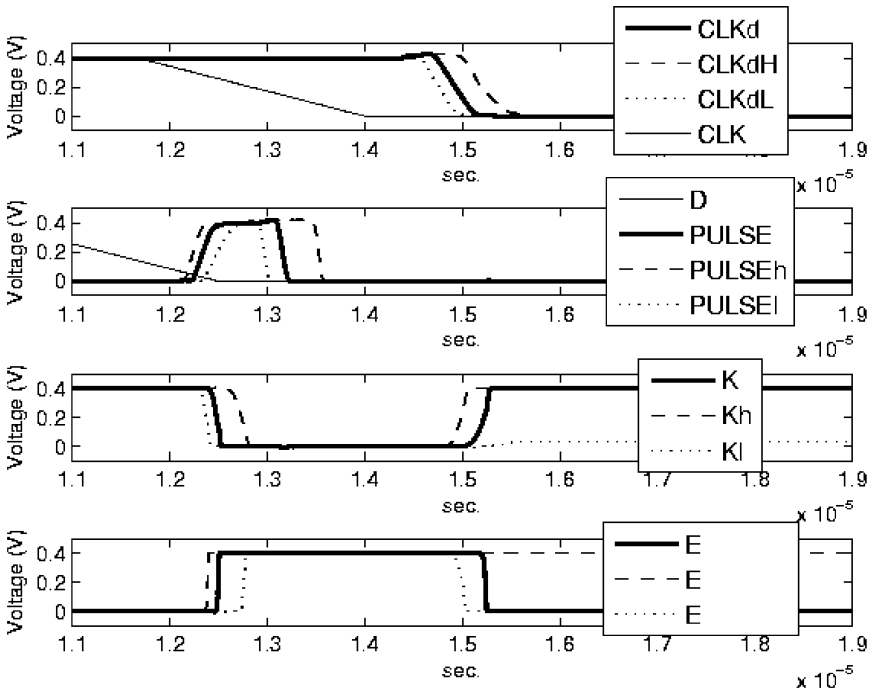


Figure 4.5: 50 point Monte-Carlo simulation with typical process corners showing node K (incorrectly) locked low during an *ERROR*-reset. The circuit for this simulation did *not* contain TG.

LATCH2. LATCH2 latches *E* high until the next cycle.

The CDC sizing is important to the accuracy of timing error detection. Inverters 3 and 4 were sized much larger than the sizing metric in Table 3.2 but provided small variation at the rising/falling edge of *CLKd*. Inverters 5, 6, and 7 were used to emulate the delay of *CLKi* and *CLKd*. This resulted in less variation at the edges of *CLKd* due to the TG connection. However, the cost of using inverters 5, 6, and 7 was larger energy consumption.

To ensure operation from V_{dd} 0.2 V to 1.2 V and to add TED functionality into LATCH1, the area of *TDTBsubI*'s layout was approximately 28 times larger than LATCH1. As a result, the average energy was an average of 15-65 times larger than LATCH1 for V_{dd} 0.2 V to 1.2 V. Additional details about the layout and energy consumption is found in Section 5.2.2.

4.2.3 Leakage

To reduce leakage, the majority of transistors L in Fig. 4.3 were sized as Long-L transistors (i.e. nominal + 10%) since long-L transistors [30] have 3x lower leakage. HVT was chosen for most transistors since it results in at least an order of magnitude decrease in leakage current per μm of NMOS (or PMOS) width. This is consistent with (3.15) which shows that as V_T goes to ∞, the leakage energy per operation goes to 0.

The majority of transistors in Fig. 4.3 use HVT to reduce leakage but a closer examination of the Keeper is needed in order to choose HVT or LVT. When *CLKd* is high and *D* transitions, a *PULSE* results at N1. Node K should be driven low, thus initiating an *ERROR*

signal. The main goal is to get *ERROR* generated as fast as possible, which is especially important for the edge case when *D* transitions at the same time *CLK* falls. The *ERROR* signal needs to be loaded into a latch or flip-flop.

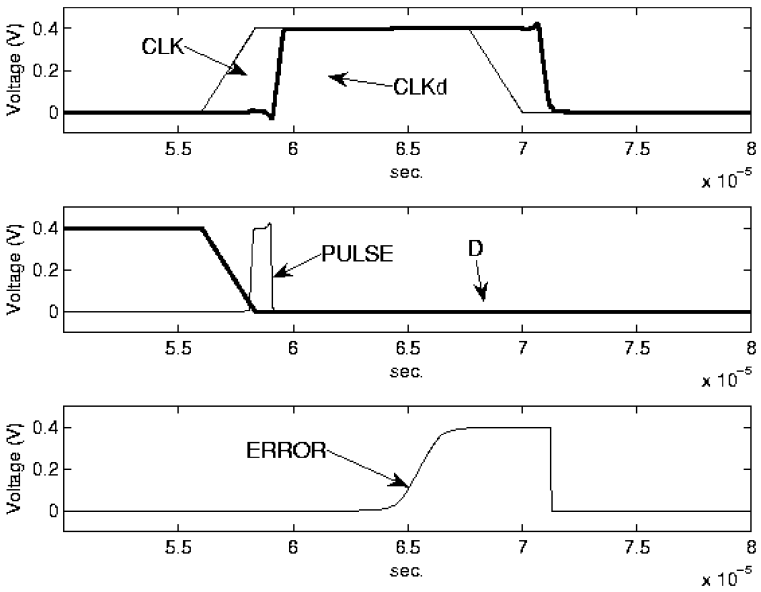


Figure 4.6: Leakage at V_{dd}=0.4 V when N1 is LVT.

In order to get *ERROR* high quickly during *CLKd* high and at the same time prevent leakage problems during *CLKd* low, two steps were taken. First, HVT was chosen for N1 and LVT for N2. This resulted in a larger current compared to as if both N1 and N2 were HVT. The reason is seen from (3.5) and (3.6); LVT devices have much larger driving strength than HVT. When *CLKd* transitions low, N1 is HVT and thus prevents leakage from node K. If N1 is an LVT device, *TDTBsubI* will not operate below 0.4 V as shown in Fig. 4.6. Every time *CLKd* transitions high, node K is (incorrectly) driven low, and consequently *ERROR* high, due to the leakage through N1. During *CLKd* high, N2 has the ability to be ON and thus N1 should be HVT to prevent leakage. The second step taken to prevent leakage problems was making P1 an HVT device since when it is off, node K may be low and leakage is not desired. Although PMOS have lower leakage than NMOS [2], the sensitivity of K requires extra leakage protection.

4.2.3.1 Leakage Detector

Although there may be leakage issues in the Keeper of Fig. 4.3, the leakage within the Keeper could be a useful tool. Understanding the amount of leakage in a digital circuit is worthwhile for some systems [31]. Using some elements of Fig. 4.3 and performing a number of modifications, a leakage detector could be constructed (Fig. 4.7). When *CLK* is low, node K's state would be maintained by the Keeper. When CLK is high, leakage through N1 would cause K's state to change (i.e. similar to Fig. 4.6 at 65 μs) and thus change E. The

leakage threshold could be set by sizing N1 and N2 accordingly.

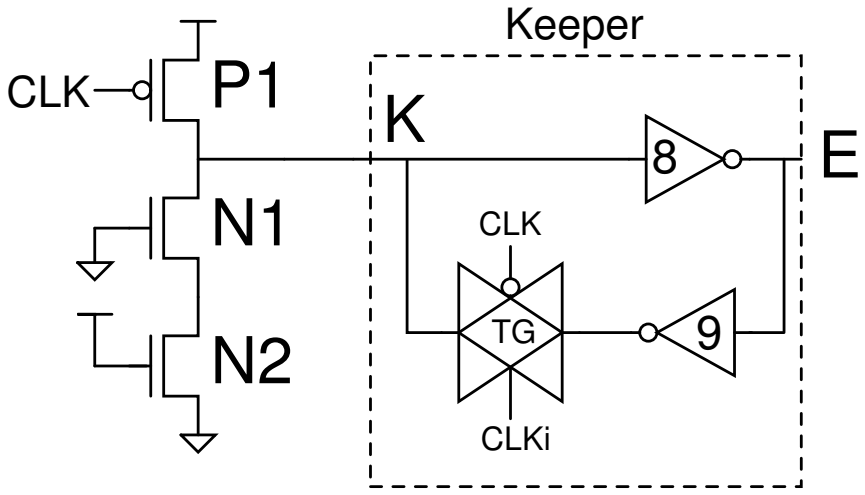


Figure 4.7: Leakage detector.

4.2.4 Design for System-level

In addition to the circuit-level issues addressed in the previous sections, the system-level needs to be considered for *TDTBsub*. At this stage in *TDTBsub* design, the most relevant system-level issue was concerned with the management of the *ERROR* signal. As shown previously in Fig. 2.4, the *ERROR* signal from each logic stage is brought to one large OR gate. Every combinational logic stage of the pipeline has a TED register which consists of N latches. If the pipeline has S number of pipeline stages, then there are N*S TED latches and N*S inputs to the OR-gate. To understand if a timing *ERROR* occurred in the pipeline, all N*S *ERROR* signals are combined into one signal using an N*S input OR-gate.

To realistically implement the logic of an N*S OR-gate, OR-gates from the standard cell-library must be used. Since the cell-library's OR-gates typically contain 2 to 3 inputs, using one large N input OR-gate is unrealistic for multi-bit systems (i.e. 32 bit and therefore N=32). Typically, combining a large number of the cell-library's 2 or 3 input OR-gates is done by creating an OR-gate tree. The timing delay of this tree needs to be considered though. If an *ERROR* occurs close to a falling *CLK* edge, the propagation delay of this signal through the OR-gate tree may be longer than the time before *CLK* transitions low. For this case, the *ERROR* would be missed for that *CLK* cycle.

To address this issue, *TDTBsub* latch has been designed so that its *ERROR* output is reset at the rising edge of *CLK*. In other words, no matter where an *ERROR* is generated under a *CLK* high, it remains high until *CLK* rises again. This ensures that timing errors which occur close the *CLK* falling edge are correctly identified provided that the time from the falling edge to the rising edge of *CLK* is longer than the OR-gate tree.

4.3 *TDTBsubII*

An explanation of the functionality of each device within *TDTBsubII* is first presented in Section 4.3.1. The sizing and circuit style used in allowing *TDTBsubII* to operate from the sub-threshold to strong inversion region is explained in Section 4.3.2. The effects of leakage on *TDTBsubII* is then given in Section 4.3.3. A more accurate derivation of $t_{r,f}$ for *D* and *CLK* is provided in Section 4.3.4. Finally, a qualitative comparison of both *TDTBsubI* and *TDTBsubII* is given in Section 4.3.5.

4.3.1 Functionality

TDTBsubII, shown in Fig. 4.8, consists of a cell-library positive edge triggered latch (LATCH), a transition detector (TD), and a reset-and-delay unit (R&D) for resetting the *ERROR* signal and to provide useful delay signals for the SN and Keeper. Fig. 4.9 shows the operation of the *TDTBsub* at V_{dd}=0.3 V. When *D* transitions under *CLKd* high, or the timing error detection window (TED window), it is considered a timing error and an *ERROR* signal is generated. The *ERROR* signal is carried high until the next TED window in order to provide robustness at the system-level as previously explained in Section 4.2.4.

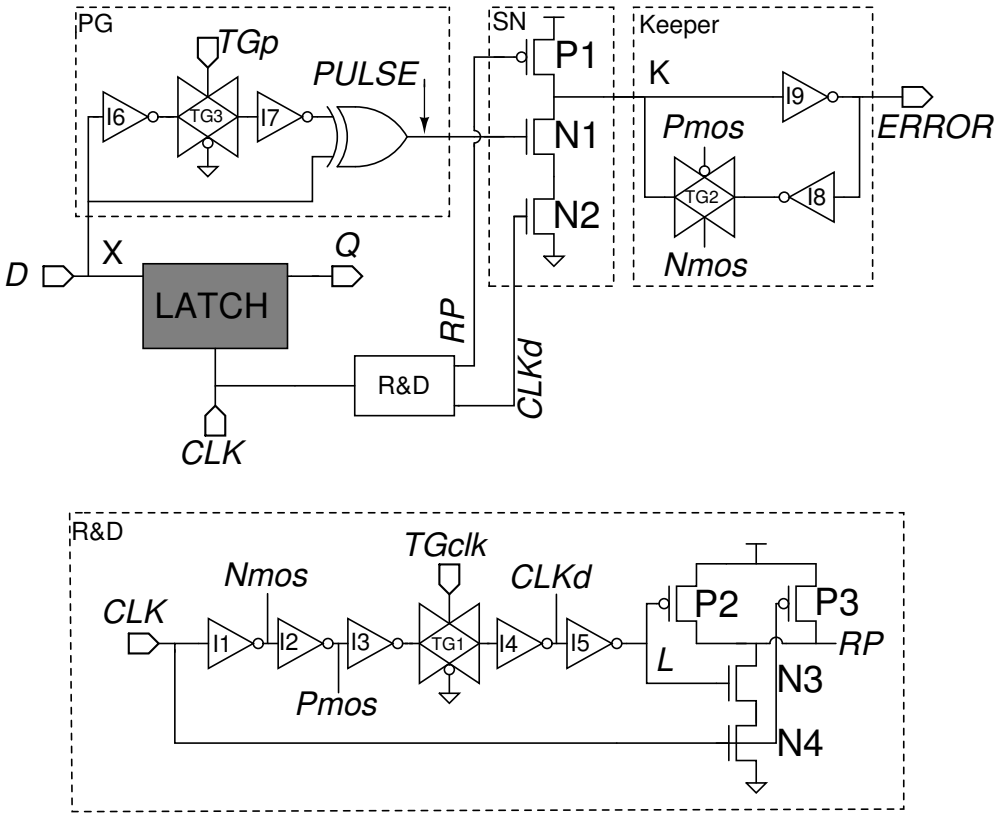


Figure 4.8: TDTBsub circuit capable of operating from V_{dd}=0.3 V to 1.2 V.

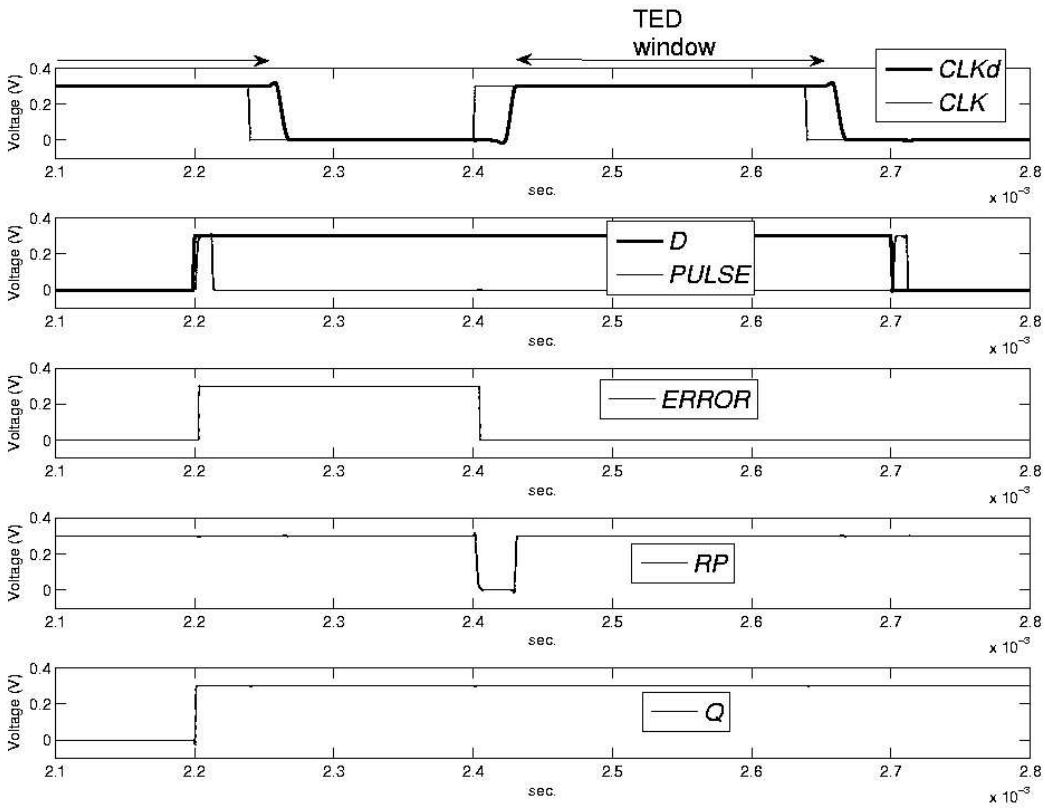


Figure 4.9: *TDTBsubII* operation at V_{dd}=0.3 V for typical process parameters. When *D* transitions under the TED window high near 2.2 ms, it is considered a timing error (*ERROR*).

The TD consists of three main components: a Keeper, a switching network (SN), and a pulse generator (PG). With the exception of TG2's control signals, the Keeper functionality is similar to *TDTBsubI*. To reduce energy consumption, TG2's control signals do not come from an emulated version of the *CLK* delay signals as in *TDTBsubI*. TG2's control signals *Nmos* and *Pmos* from the R&D are important to the operation of the Keeper. When *CLK* is low, the TG is *OFF*, which makes it easier for node K to be driven low when a *PULSE* arrives at N1. As *CLK* transitions high, *Nmos* and *Pmos* switch the TG ON again and incorrect timing errors due to leakage are prevented.

SN is synchronized with *CLKd* and RP to pull node K low if *D* transitions at the same time *CLKd* is high and reset *ERROR* at the next TED window, respectively. The operation of the SN highlights the most fundamental operational differences between *TDTBsubI* and *TDTBsubII*. First, SN turns P1 ON *only* when *CLK* transitions low to high. This means that P1 is required to switch only I9 during an *ERROR* reset. Secondly, when K is required to stay high, TG2 turns the Keeper ON thus preventing leakage currents from switching K. The altered timing control of P1 in *TDTBsubII* allows for the removal of LATCH2 from *TDTBsubI*.

PG uses inverters I6 and I7, a transmission gate (TG), and an XOR to generate a short voltage pulse (*PULSE*) as shown in Fig. 4.9 each time the data (*D*) transitions. The inverters and TG determine the duration of the XOR's *PULSE*. Similarly to *TDTBsubI*, a *PULSE* with

small delay and small duration is desirable. Adding a control voltage to TG3 provides post-manufacturing tuning capability [15] of the *PULSE* duration. To operate *TDTBsubII*, tuning was not required during simulation (i.e. TG3 was set to V_{dd} for all simulations).

The R&D unit provides a delayed version of *CLK*, generates signals for the Keeper, and gives a signal to reset the *ERROR*. A delayed version of *CLK*, or *CLKd*, is essential to ensure that transitions of *D* near the rising edge of *CLK* do not incorrectly generate *ERROR*s. To prevent legitimate *D* transitions from triggering an *ERROR*, the *CLK*-*CLKd* delay at the *CLK* rising edge has two requirements. First, *CLK*-*CLKd* delay must be longer than the duration of *PULSE*. Secondly, *CLK*-*CLKd* should be larger than the maximum CLK-Q delay of LATCH. The duration of the *CLK*-*CLKd* delay at the falling edge of *CLK* is insignificant assuming the duty cycle (*d*) of *CLK* can be adjusted at the system-level.

RP is an important signal within the R&D used to reset the *ERROR* for a new TED window. During the time that *CLK* is low, P3 is turned ON and keeps *RP* high. When *CLK* transitions from low to high, P3 is shut OFF and N4 is turned ON. N3 remains ON for the duration of CLK-L delay and as a results pulls *RP* low. This causes P1 to turn ON and flip the state of the Keeper. Node K is driven high and *ERROR* is reset low for the next TED window. This event happens only when *CLK* goes from low to high. Once the duration of CLK-L delay is reached, P2 turns ON and N3 turns OFF thus forcing *RP* high again. The duration of the *RP* low is dependent on the CLK-L delay.

Similarly to TG3 in PG, TG1 within the R&D unit provides post-manufacturing tuning capability. Since the *CLK*-*CLKd* delay changes with variation, the use of tuning will allow for improved accuracy in *TDTBsubII*. TG1 was set to V_{dd} for all simulations.

4.3.2 Sizing & Circuit Style

The sizing and circuit style of *TDTBsubII* was examined for each element in the TD. The TD consists of three devices, including the PG, SN, and Keeper. The PG consists of two inverters, a TG, and an XOR which are used to generate the *PULSE* signal. The size used for each device is shown in Table 4.4. The challenge was finding a *PULSE* width with a small duration that did not change its shape during Monte-Carlo simulations within the sub-threshold region. Inverters I6, I7, and TG were first sized by using Table 3.2. The inverter sizes and TG were further tuned until both rising and falling transitions of *D* produced similar *PULSE* results. Small incremental changes to the size of I6 and I7 were made due the sensitivity to power consumption with respect to I6 and I7. The XOR used was the same as in *TDTBsubI*.

The SN proved to be the most challenging in sizing for the sub-threshold region operation. Both the threshold types and sizes of transistors were examined. The threshold types are discussed in the next section. Transistors N1 and N2 were sized larger than W=5.33*W_{nMIN} to reduce variability. For a 50-point Monte-Carlo *without* TG2, P1 is unable to drive *K* high as shown in Fig. 4.10. The stronger (NMOS) of I8 works against the weaker P1 (PMOS)

Table 4.4: Device sizing for *TDTBsubII*

Device	PMOS W/L	NMOS W/L	Device	PMOS W/L	NMOS W/L
I1-I4	1.5/0.25	0.8/0.30	N2	N/A	3.0/0.08
I5	0.54/0.08	0.36/0.08	N3=N4	N/A	0.6/0.08
I6	0.6/0.12	0.4/0.22	TG1	0.36/0.25	0.36/0.25
I7	1.35/0.12	0.6/0.18	TG2	1/0.08	1/0.08
I8	1.0/0.08	0.4/0.08	TG3	0.4/0.08	0.4/0.08
I9	0.6/0.08	0.4/0.08	P1	2.0/0.08	N/A
N1	N/A	2.0/0.08	P2=P3	0.5/0.08	N/A

and wins to keep K (incorrectly) low at 2 ms. Making P1 extremely large is an option, but this results in increased leakage and unwanted delay to *CLKd*. The best solution is to size P1 according to Table 3.2 and close the feedback path to I8 by use of a TG.

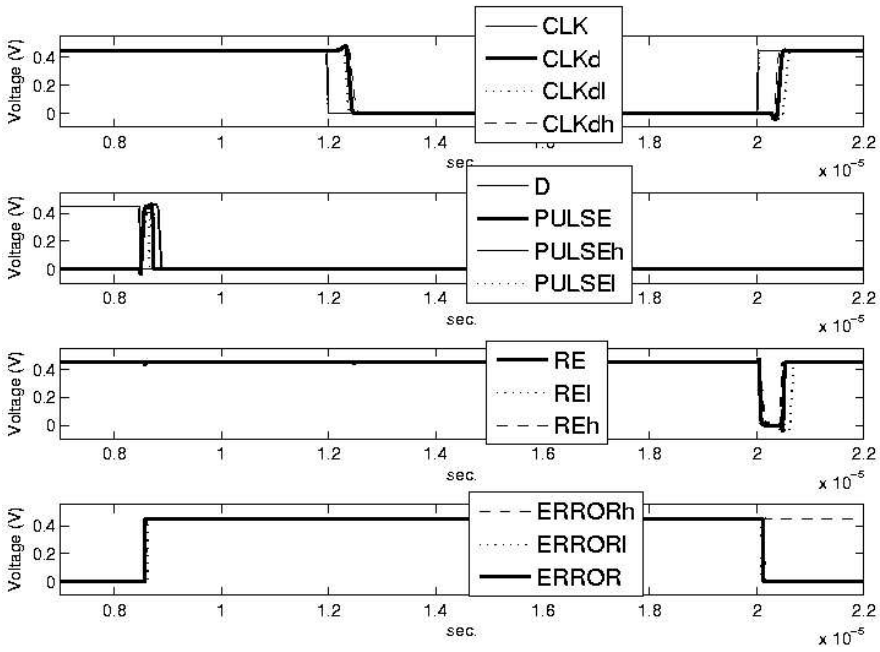


Figure 4.10: 50 point Monte-Carlo of *TDTBsubII* without TG2 at V_{dd}=0.45 V.

The sizing of the Keeper was important with respect to leakage. When *CLK* is low, a transition of *D* should not cause *ERROR*s. If the leakage through N2 is large and the Keeper is not strong enough to keep K high, an *ERROR* may be inadvertently generated when *D* transitions under a *CLK* low. As a result of this issue, the size of TG should be large (i.e. width of NMOS and PMOS set to 1 μm) and the width of PMOS of I8 large (i.e. 1 μm). Making these both large allows for K to remain high even for large values of leakage through the SN. Inverter I9 was sized using the recommendations in Section 4.1 (i.e. K=1.5).

CLKd and the duration of *RP*'s low pulse is dependent on the size of components in the R&D including I1-I5, the TG1, P2, P3, N3, and N4. The inverters I1-I5 were also sized with K=1.5. Variation of *CLKd* should be minimal to improve the accuracy of *TDTBsub* and thus the sizes of the inverters I1-I5 and TG1 were larger than required by Table 3.2. The size of

N3, N4, P2, and P3 were also sized larger than Table 3.2. The size of these transistors affects the duration of *RP*'s low pulse.

To ensure operation from V_{dd} 0.3 V to 1.2 V and add TED functionality to LATCH, the area of *TDTBsubII*'s layout was 11 times larger than LATCH. As a result, the average energy was 13.8 times larger than LATCH averaged for V_{dd} 0.3 V to 1.2 V. Additional details about the layout and energy consumption is found in Section 5.3.2.

4.3.3 Leakage

The dominant leakage mechanism in *TDTBsubII* was a combination of *Ig* and I_{CH} (3.1.3). To address leakage, the lengths (L) of all transistors were first adjusted. Next, a number of steps were taken to choose the correct transistor threshold voltages. To reduce leakage, the majority of transistors L were sized as Long-L transistors (i.e. nominal + 10%).

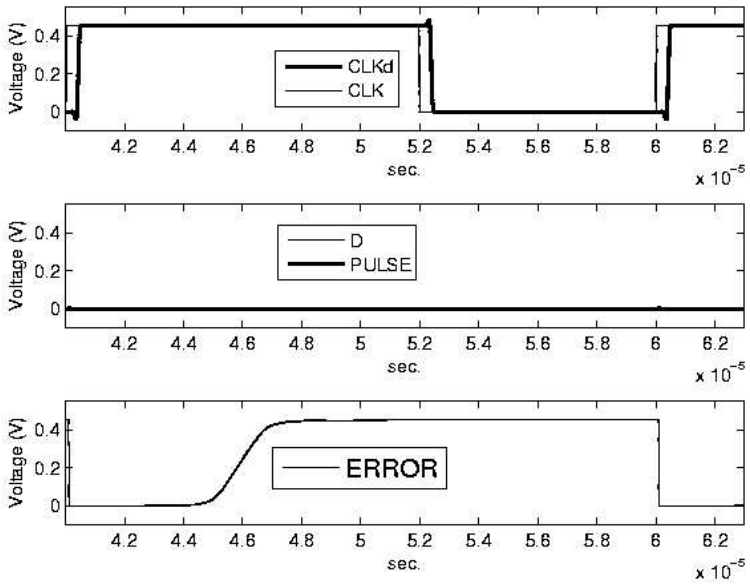


Figure 4.11: *TDTBsubII* at typical process corners with N1 and N2 sized as LVT. Leakage through N1 causes the signal at K to go low and thus signal *ERROR* to go high. This situation is prevented by making N1 HVT and N2 SVT in addition to sizing the TG2 and I8 of the Keeper large.

Compared to *TDTBsubI*, more steps were required to rapidly get *ERROR* high during *CLKd* high and at the same time prevent leakage problems during *CLKd* low. First, N1 was set as HVT and N2 as SVT. This resulted in a larger current when driving K low for a timing error compared to as if both N1 and N2 were HVT. If N1 is an LVT device, *TDTBsubII* will not operate below 0.45 V as shown in Fig. 4.11. Each time *CLKd* transitions high, node K is (incorrectly) driven low due to the leakage through N1. During *CLKd* high, N2 has the ability to be ON and thus N1 should be HVT to prevent leakage. The second step taken was making P1 an HVT device as in *TDTBsubI*. Thirdly, TG2 was turned ON when *CLK* was low. Thus, if there was leakage in the SN, it was combated by a large TG2 and I8. Finally,

I9 should have NMOS and PMOS that are LVT. This provides a lower delay as K transitions low and thus *ERROR* is generated more quickly.

4.3.4 Rise and Fall Times

The $t_{r,f}$ of the *CLK* and *D* is important to the performance of *TDTBsub*. Changing the $t_{r,f}$ of *CLK* changes the size of the TED window while a change in $t_{r,f}$ of *D* translates into a new duration of *PULSE*. Considering these issues, it is crucial to pick reasonable values of $t_{r,f}$ times for both *CLK* and *D* for simulations of V_{dd} from 0.2 V to 1.2 V.

A 5-stage ring oscillator was constructed to attain more accurate values of $t_{r,f}$ times for each V_{dd} from 0.2 V to 1.2 V. The oscillator is considered a standard for delay measurements [17]; delay is a function of $t_{r,f}$ times. As shown in Fig. 4.12, the $t_{r,f}$ times increased as V_{dd} is lowered. As V_{dd} is lowered into the sub-threshold region, the $t_{r,f}$ becomes exponentially dependent on V_{dd} as predicted by (3.8). Note that the $t_{r,f}$ times are defined as the time interval between the signal crossing 20% of V_{dd} and 80% of V_{dd}.

The $t_{r,f}$ time of a signal is largely determined by the capacitive load presented to it and the strength of the driving gate. For *TDTBsubII*, the capacitive load at node X in Fig. 4.8 is affected by the size of I6, LATCH, and the XOR. To account for their capacitance, the output values from Fig. 4.12 were multiplied by 2 since the capacitance simulated at node X was about 2 times the input capacitance of the oscillator.

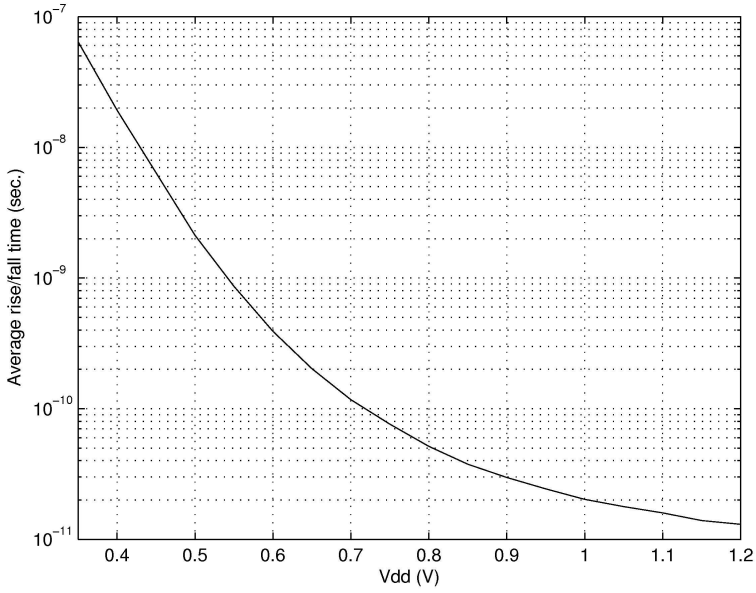


Figure 4.12: Rise/fall times as a function of V_{dd} for a 5-stage ring oscillator.

4.3.5 Comparison to *TDTBsubI*

TDTBsubII was a significant improvement over *TDTBsubI* due to changes in *d*, removal of emulation inverters 5,6, and 7, the addition of the R&D unit, and more accurate values of $t_{r,f}$

for simulation. At the system-level, *d* can be easily adjusted, therefore, the *d* was lengthened in *TDTBsubII* to more accurately account for cases when *D* transitions when *CLK* falls. The emulation inverters of *TDTBsubI* were removed due their large power consumption and small effect on reducing variation. The R&D unit provided a more energy-efficient method of keeping the timing error signal high until *CLK* transitioned low-to-high. Using a ring oscillator to understand the $t_{r,f}$ as V_{dd} changed allowed for more accurate sizing of inverters in the PG. With the new $t_{r,f}$ times and the realization that small adjustments of $t_{r,f}$ could be made at the system-level, the power-hungry inverter 0 was removed. Overall, *TDTBsubII* has a number of improvements (e.g. energy consumption, robustness, etc.) over *TDTBsubI*.

4.4 System-level Test Circuit

A system-level test circuit (*SystemTest1*) was designed to test the operation of *TDTBsubI* at a system-level (Fig. 4.13). *SystemTest1* used *TDTBsubI* latches because *TDTBsubII* was not ready by the chip deadline for this Master's thesis. Although *TDTBsubII* shows improvements over *TDTBsubI*, the functionality of both versions is similar within *SystemTest1*. *SystemTest1* consists of input (shiftIN) and output (shiftO) shift registers operating at Vddh=1.2 V, *TDTBsubI* latches and an adder both operating at V_{dd} (0.2 V to 1.2 V), and level-shifters (levelS). Amongst these components, the *TDTBsubI* and the level-shifters were not taken from the cell-library but instead designed in this thesis work. Since the design of *TDTBsubI* has already been presented, only the level-shifter design is given here.

A level-shifter was needed to shift the level of voltage used within the combinational logic of *SystemTest1* in Fig. 4.13. The combinational logic voltage V_{dd} can vary from 0.25 V to 1.2 V while the input and output register V_{ddh} is always at 1.2 V. Therefore, a level-shifter was designed to shift any logic at V_{dd} back to 1.2 V at the output shift registers (Fig. 4.15). As shown in Fig. 4.13, 60 level-shifters (i.e. levelS) are placed after the combinational logic.

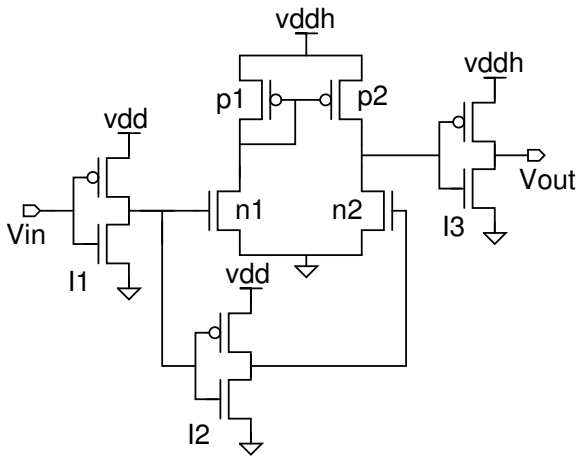


Figure 4.14: Level-shifter.

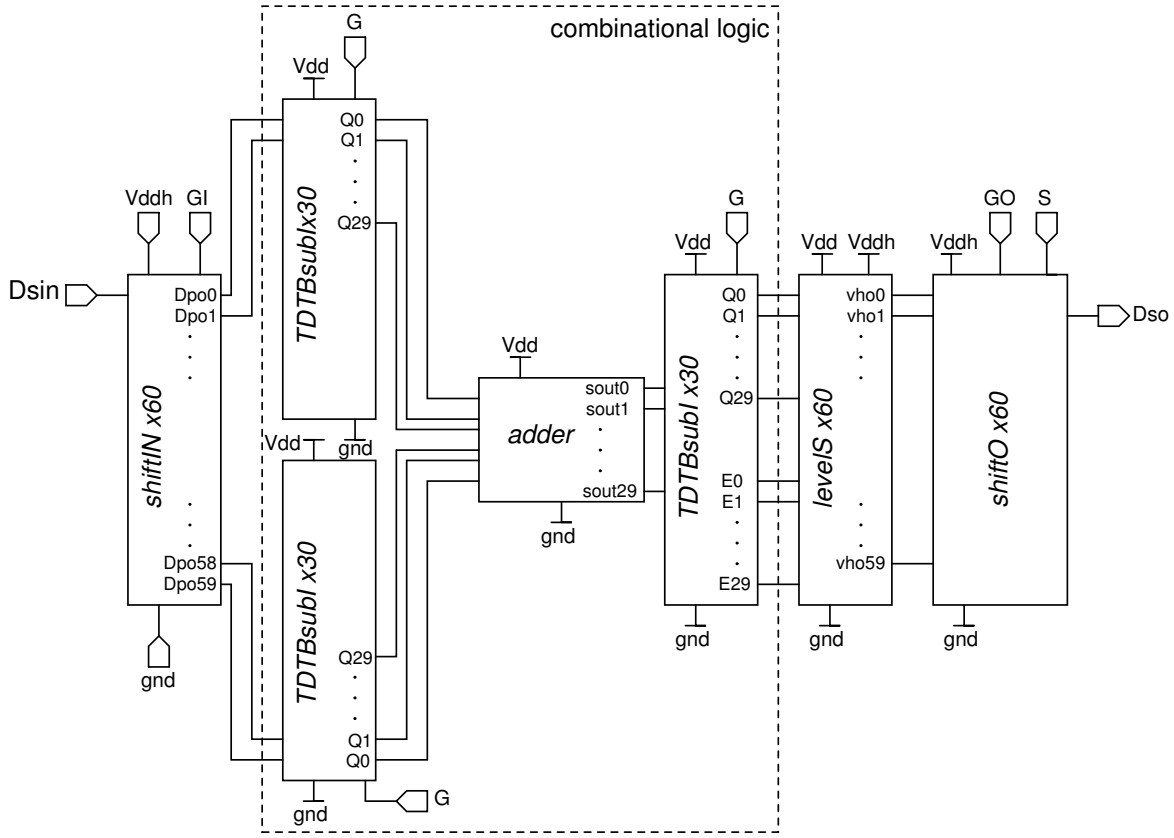


Figure 4.13: *SystemTest1*.

The level-shifter circuit is similar to a conventional level-shifter [32] which is designed for robustness in the sub-threshold region. However, it was not designed with concern for energy consumption. The level-shifter is a differential amplifier with a constant current mirror load. This load has increased stability compared to a cross-coupled load for the sub-threshold region. It is well suited for a wide voltage range but suffers from increased energy consumption [32]. The inverters I1, I2, and I3 were sized according to Table 3.2 and using K=1.5 as recommended in Section 4.1. PMOS devices p1 and p2 were sized minimally while the NMOS devices n1 and n2 were required to have a width of 6 μm to function into the sub-threshold region. A 1000 point Monte-Carlo simulation at typical process corners shows that the level-shifter works correctly down to 0.2 V (Fig. 4.15). As a result of its robustness, it was used in *SystemTest1*.

4.5 Layout

Following the design of the *TDTBsub* (Section 4.2 and 4.3) and *SystemTest1* (Section 4.4), the layout of a single *TDTBsubI* latch, the level-shifter, and *SystemTest1* were designed in 65 nm CMOS. The layout of *TDTBsubI* is shown in Fig. 4.16. Due to the long length of the layout, only a few components are shown; see Appendix A for the full layout. The long length is a result of the large *CLK* delay components (i.e. the CDC of Fig. 4.3).

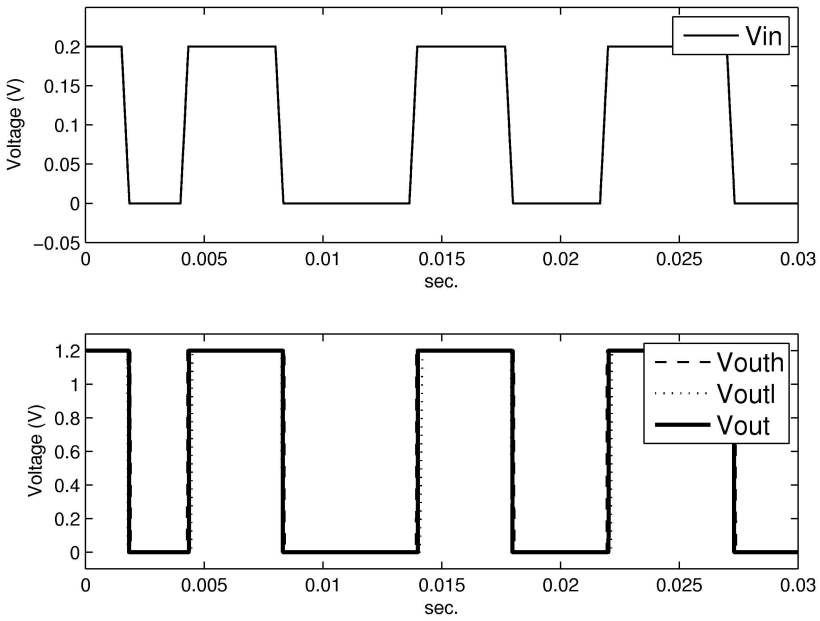


Figure 4.15: Level-shifter for a 1000-point Monte-Carlo simulation.

Approximately 70% of the area of *TDTBsubI* is used for the CDC. The data (*D*) and *CLK* signals are brought to the left of the layout into LATCH1. Output signals *ERROR* and *Q* are at the far right side of the layout; these are not visible from Fig. 4.16. After the *TDTBsubI* layout was complete, the level-shifter layout was designed (see Appendix A). Using both the *TDTBsubI* and level-shifter layout, *SystemTest1* was built (Fig. 4.17).

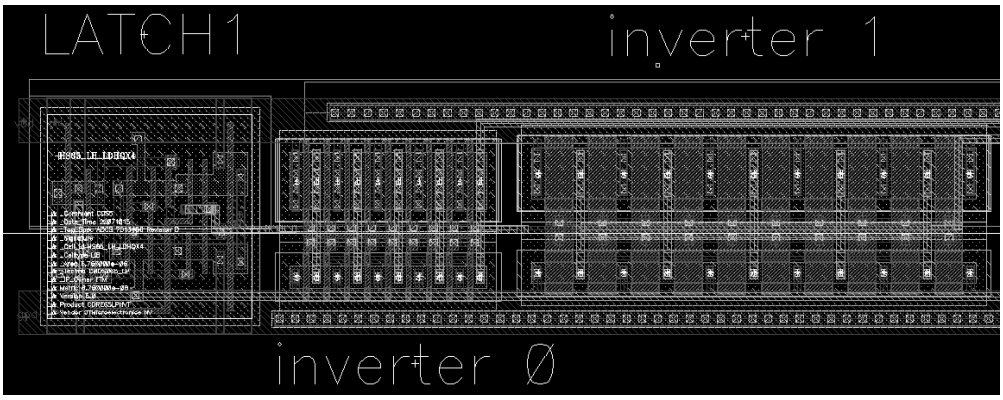


Figure 4.16: *TDTBsubI* layout.

Once the *TDTBsubI* latch and *SystemTest1* were complete, a top-level (i.e. last layout before manufacturing) layout was designed (Fig. 4.18). In addition to a single *TDTBsubI* latch and *SystemTest1*, the top-level layout included a pad ring. It provides a location to wire bond the chip package to, connects the top layer metal within the chip (i.e. metal 6) to the wire bonding pad, and is able to the protect the circuit from unwanted electrostatic discharge (ESD). To operate the top-level correctly, the pad ring must be supplied with 1.8 V voltage supply. This supply allows for a pull-down or pull-up network to drive the pad output. Two

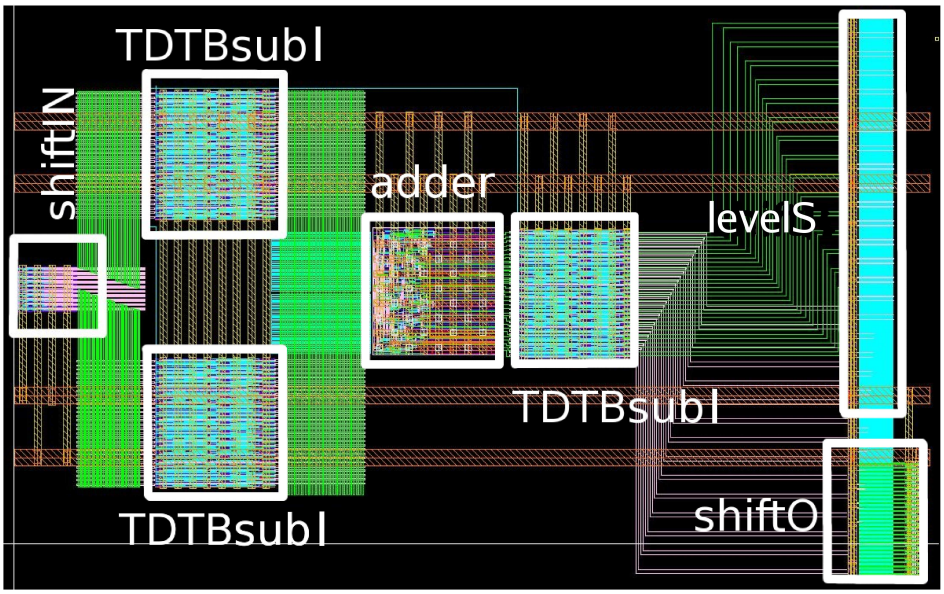


Figure 4.17: *SystemTest1* layout.

enable signals were connected within the pad ring to activate these networks.

Connecting the pad rings to the *TDTBsubI* latch and *SystemTest1* was also required in the top-level layout. Due to the limited number of available pad rings, some connections had to be shared. This causes limitations in performance (i.e. energy consumption, speed, and functionality) measurements. For example, the Vdd of *TDTBsubI* latch is connected to *SystemTest1* Vdd. Therefore, only the speed and functionality of *TDTBsubI* latch can be accurately measured since energy consumption is adversely affected by the much larger leakage of *SystemTest1*.

A photomicrograph of the post-manufactured top-level layout is shown in Fig. 4.19. Although the view is the same as in Fig. 4.18, there is an important difference between Fig. 4.18 and 4.19. A metal 6 layer should be visible between the wire bonding pads and on-chip components (i.e. *TDTBsub* latch and *SystemTest1*) in Fig. 4.19. Additionally, an emblem of the manufacturer's name should be written with the metal 6 layer. The metal 6 layer was not placed during the manufacturing process and was beyond the control of the design in this thesis. As a result, the connection to the *TDTB-subI* latch and *SystemTest1* is not possible through the wire bonding pads. Therefore, measurements of both components is not possible due to this manufacturing process error.

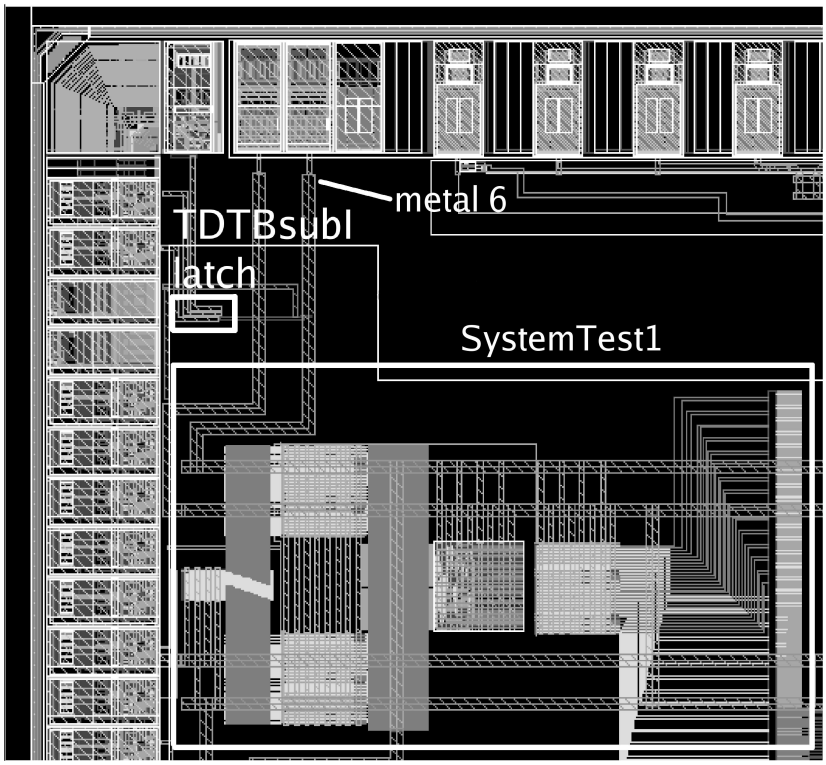


Figure 4.18: Top-level layout.

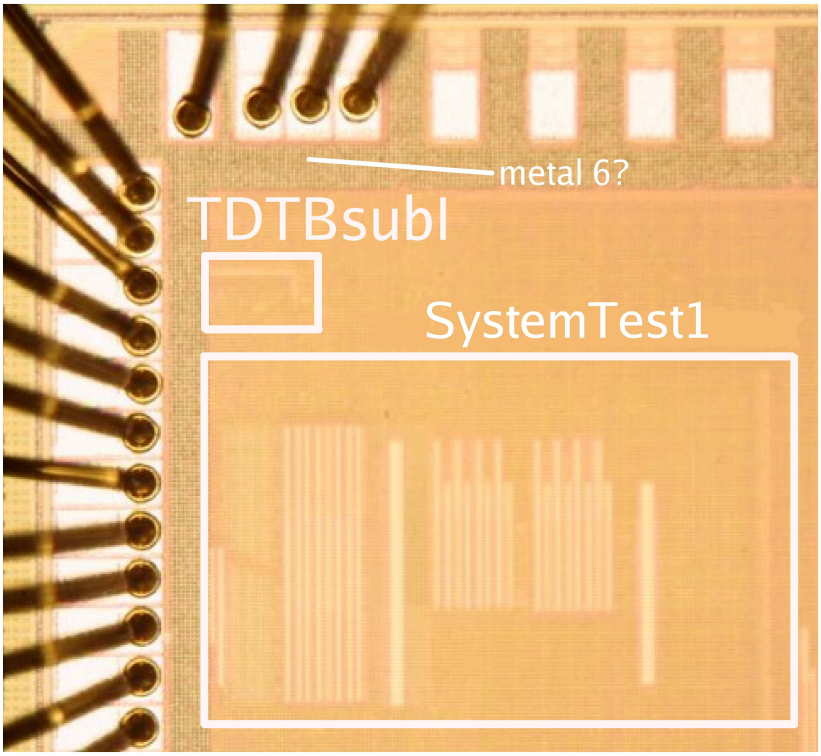


Figure 4.19: Photomicrograph of the implemented top-level layout.

4.6 Measurement System

To measure the performance and functionality of *TDTBsubI* and *SystemTest1*, a measurement system was designed (Fig. 4.20). Although the system was not utilized due to the manufacturing error previously mentioned, it is ready to use upon completion of the future manufactured chip. The measurement system consists of the Tektronix TLA720 logic analyzer, a PCB (Fig. 4.21), and an oscilloscope. Within the logic analyzer, basic binary *D* input test signals are written and sent to the PCB. For example, *D* was designed in the logic analyzer to transition under *CLK* high and *CLK* low. The PCB was designed to provide low noise when measuring *TDTBsubI* and the test circuit. To reduce noise, capacitors were placed between each supply signal and ground. Additionally, ground and supply planes were used to reduce noise. The oscilloscope is used to the measure the output signals from the PCB.

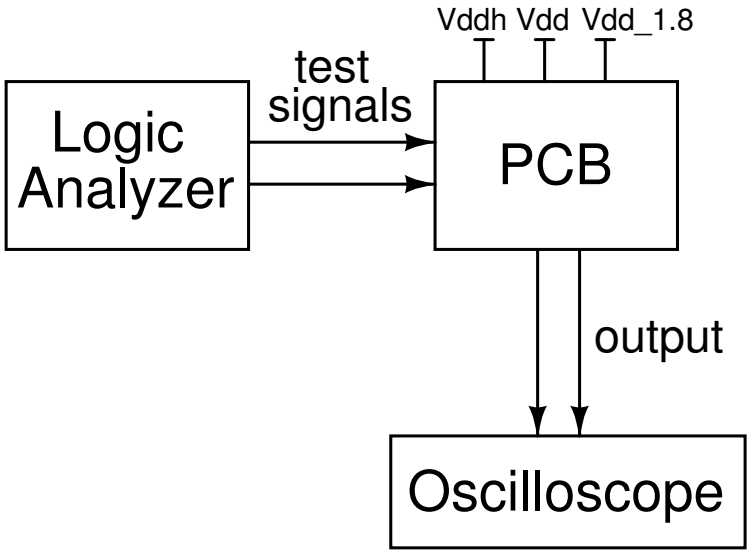


Figure 4.20: Measurement system.

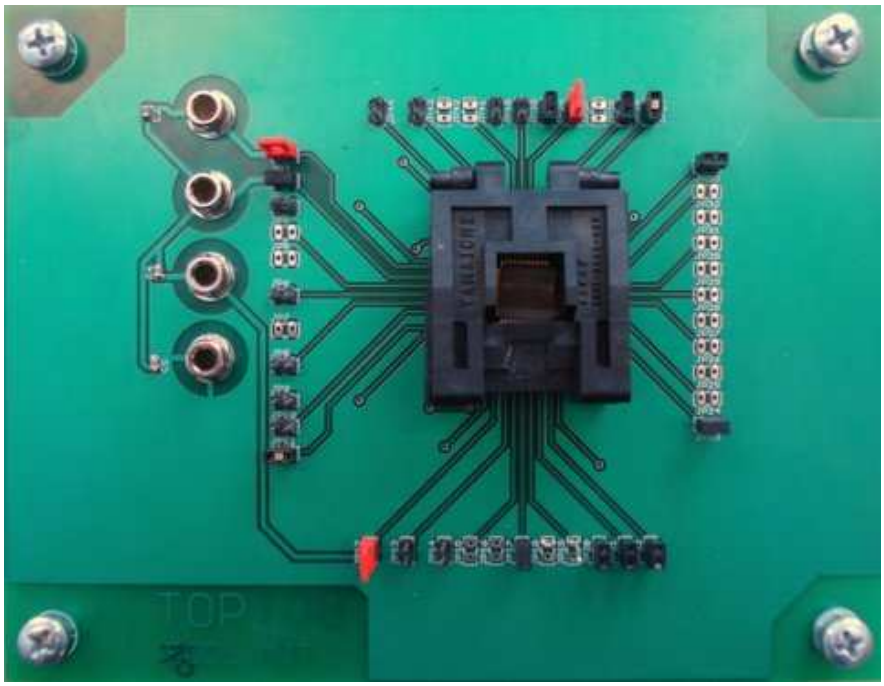


Figure 4.21: PCB.

Chapter 5

Simulation Results

In this chapter, the simulation results of *TDTBsub* are presented. During initial stages in simulating *TDTBsub*, it was decided that a testing plan was needed to standardize the simulation process. Thus, a testing plan was developed and is described in Section 5.1; it was used throughout the simulations in this chapter and also the design process in Chapter 4. Next, the results of simulations are shown in Section 5.2 and Section 5.3 to understand the fundamental characteristics of *TDTBsub* including operating frequency, energy consumption, and global and local variation impacts on functionality. A comparison between both *TDTBsub* versions is presented in Section 5.4. Finally, simulation results of a system-level test circuit are given in Section 5.5.

5.1 *TDTBsub* Testing Plan

To ensure the correct operation of *TDTBsub* from sub-threshold to strong inversion region and provide standardization in the simulations, a testing plan was created. The testing plan includes two test cases:*D* transitions and *CLKd* delay constraints. The three requirements for *D* transitions (with corresponding diagram in Fig. 5.1) were:

1. Latch setup times: *D* is transitioned at the latch setup time for a falling edge of *CLKd*. An *ERROR* should result and *D* should be latched.

2. Edge transitions: For the rising edge of *CLKd*, there should be not be an *ERROR* if *D* transitions at the same time as *CLKd*. For the falling edge of *CLKd*, an *ERROR* should be generated if *D* transitions at the same time as *CLKd*.

3. *CLKd* low: If *D* transitions when *CLKd* is low, an *ERROR* should not be generated.

In addition to the test cases above, a *CLKd* delay constraint was required within the testing plan: the delay from the falling edge of *CLK* to the falling edge of *CLKd* (t_{dc}) should be less than 10% of the period of *CLK* (T_{CLK}) as shown in Fig. 5.2. Although this constraint limits the frequency performance, it ensures that t_{dc} does not become too large of a percentage of T_{CLK}. As t_{dc} delay grows, it approaches the next rising edge of *CLK* and no longer

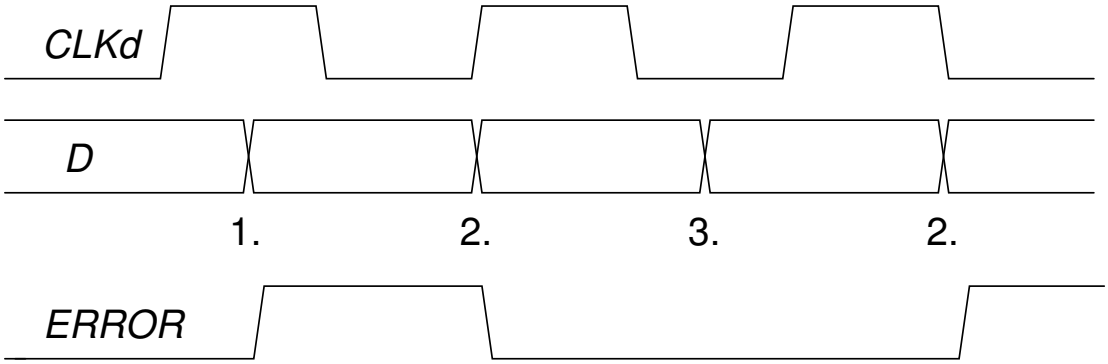


Figure 5.1: Timing diagram showing the three test cases for *D* transitions. When *D* transitions under *CLKd* high, which means *D* is arriving too late, an *ERROR* signal results. No *ERROR* signal results when *D* transitions at a *CLKd* low or when *D* and *CLKd* transition at the same time.

allows correct timing error detection with regards to the real *CLK* signal. The constraint provides a reasonable distance that *CLKd* can be delayed from *CLK*.

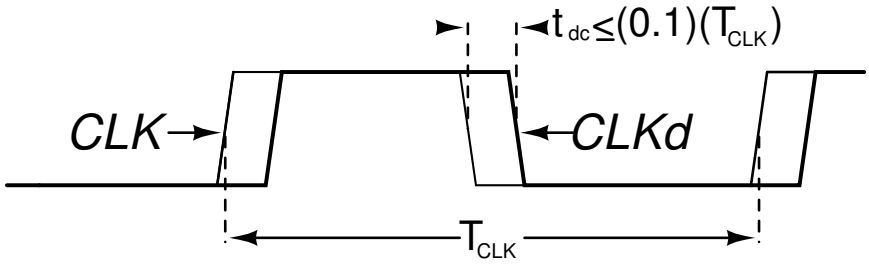


Figure 5.2: *CLKd* delay constraint.

5.2 *TDTBsubI*

Simulations of *TDTBsubI* were performed to understand the operating frequency, energy per operation, and functionality with variations. The testing plan of Section 5.1 was applied to each simulation. The $t_{r,f}$ used for *D* and *CLK* within each simulation was calculated as 5% of the *CLK* period. The *CLK* period was determined by the delay constraint in Fig. 5.2. *TDTBsubI* is capable of operation from 0.2 V to 1.2 V.

5.2.1 Operating Frequency

A simulation to understand the operating frequency of *TDTBsubI* was performed using the TT global process corner from 0.1 V to 1.2 V and a sampling resolution of 0.1 V. The results are shown in Fig. 5.3. Below $V_{dd} \approx 0.4$ V, or sub-threshold region operation, the curve shows an exponential dependence on V_{dd}. From 0.4 V to 1.0 V, the curve is quadratic. As

the voltage is increased from 1.0 V to 1.2 V, the result is a linear curve. This closely follows the curves previously predicted by the delay equations (3.5) and (3.6) in Section 3.1.1. A maximum frequency of 143 MHz was found at V_{dd}=1.2 V while a minimum frequency of 250 Hz resulted at V_{dd}=0.2 V. Below 0.2 V, the incorrection operation results due to increased leakage.

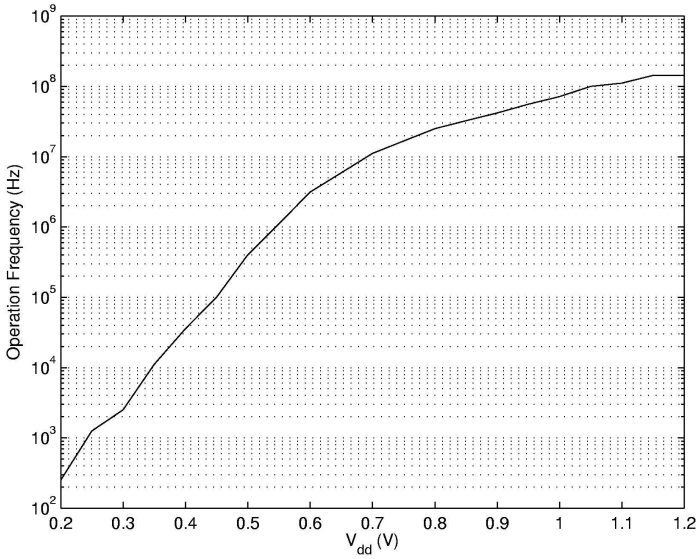


Figure 5.3: Operating frequency as V_{dd} is swept.

5.2.2 Energy Consumption

Before presenting energy consumption simulation results, it is useful to present some definitions. The total energy consumption (E_{TOT}) can be broken into two parts as explained in Section 3.3: leakage energy (E_{LEAK}) and switching energy (E_{SW}). Due to the difficulty in finding and isolating the E_{LEAK} and E_{SW}, only E_{TOT} was determined for both LATCH1 and *TDTBsubI*. More specifically, the total average energy per operation was found for both LATCH1 (E_{L1op}) and *TDTBsubI* (E_{TOTop}). An operation was considered as *D* transitioning under *CLK* high for one period. A total of nine transitions were constructed using the testing plan in Section 5.1. By integrating the total power from V_{dd} over nine periods and dividing this result by nine, E_{TOTop} and E_{L1op} were found.

To understand the amount of energy overhead due to adding TED capability to LATCH1, the energy consumption of LATCH1 (see Fig. 4.3) from *TDTBsubI* was first simulated. This allowed for a comparison to the (TED capable) *TDTBsubI* latch. The energy consumption of *TDTBsubI* was then simulated.

Fig. 5.4 shows the total average energy per operation of LATCH1 (E_{L1op}) as V_{dd} was swept from 0.2 V to 1.2 V. For most V_{dd}, the switching energy was dominant and closely followed (3.13). When V_{dd} is about 0.3 V, however, the leakage energy begins to increase

significantly thus adding to the total energy. This is because the propagation delay in the sub-threshold region increases exponentially and the leakage energy is integrated over a longer period of time. In other words, T_{op} (3.15), or the time to complete an operation, becomes exponentially long. At V_{dd}=0.3 V, E_{L1op} starts to increase again due to the increased leakage energy. The MEP is found at V_{dd}=0.4 V and is mathematically represented as the point at which the slopes of the switching energy and leakage energy are equal in magnitude and opposite in sign [2].

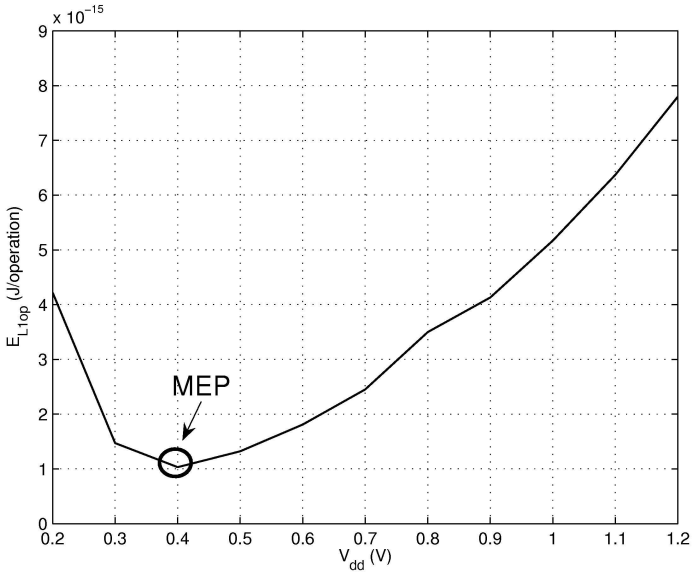


Figure 5.4: Total average energy consumption per operation (E_{L1op}) for LATCH1. The MEP is at 0.4 V.

Next, the energy consumption of the *TDTBsubI* latch was simulated. Fig. 5.5 shows the total average energy per operation (E_{TOTop}) as V_{dd} is swept from 0.2 V to 1.2 V of *TDTBsubI*. The MEP of the *TDTBsubI* latch is much smaller than LATCH1. This indicates that the total average leakage energy per operation (E_{LEAKop}) is a smaller percentage of E_{TOTop} for *TDTBsubI* than LATCH1. This is mostly due to the sizing of all L in *TDTBsubI* greater than or equal to 0.08 μm. Long-L transistors [30] have 3x lower leakage but are slower. All L in LATCH1 are sized to the minimum width of 0.06 μm.

From Fig. 5.4 and Fig. 5.5, it can be calculated that *TDTBsubI* consumes about 15 and 66 times more E_{TOTop} than LATCH1 at 0.25 V and 1.2 V, respectively. The E_{TOTop} averaged over all V_{dd} shows that *TDTBsubI* consumes about 45 times more energy than LATCH1.

5.2.3 Functionality with Variations

As previously mentioned in Section 4.2.2, careful consideration should be taken to account for global and local variations at low voltages and deep sub-micron technologies. Global variations can be applied by using a global process corner for each simulation. More importantly though, is the consideration of local variations which have become more dominant

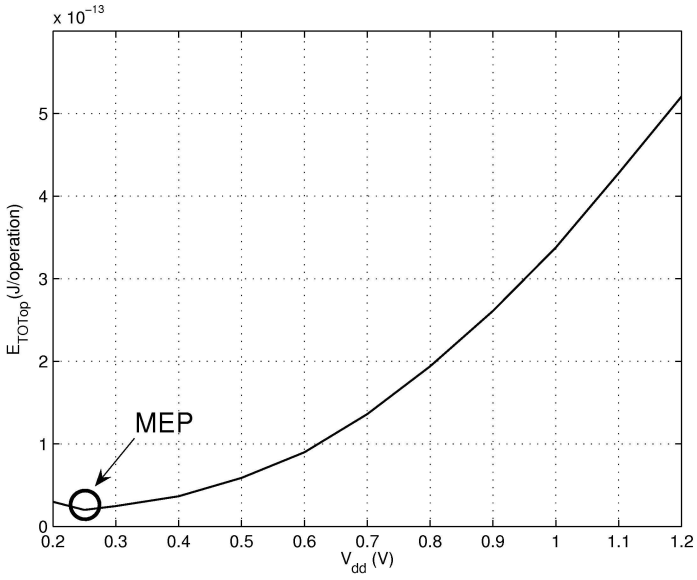


Figure 5.5: Total average energy consumption per operation (E_{TOTop}) for a *TDTBsubI* latch with a MEP at 0.25 V.

than global [26]. The local variations are understood by applying Monte-Carlo analysis [2]. Monte-Carlo analysis allows for a circuit to be simulated over a wide range of randomly chosen device parameters. For the simulations in Chapter 4 and 5, it was used to understand local variations with respect to signals at important nodes (i.e. K in Fig. 4.3).

To account for local variations in *TDTBsubI*, a 1000 point Monte-Carlo simulation was performed at the TT global process corner from V_{dd} of 0.2 V to 1.2 V. The results of the simulation at V_{dd}=0.25 V are displayed in Fig. 5.6. The results of the simulation show the average, maximum (h), and minimum output (l) values. For example, the minimum output value of *CLKd* is CLKdl. A transition of *D* under the time when *CLKd*, *CLKdl*, or *CLKdh* are high, indicates that the *D* is arriving too late and an *ERROR* signal results. Similarly, a transition of *D* before *CLKd*, *CLKdl*, or *CLKdh* are high indicates that the *D* is arriving on time and an *ERROR* signal is not applied. All three test cases for *D* transitions in Section 5.1 passed with the exception of a small number of *CLK* falling edge transitions near 0.5 V. This is most likely due to the discontinuity in the BSIM4 model which was used for all simulations of *TDTBsub*. For more information on the BSIM4 model see [33].

In addition to local variations, global variations were briefly examined for *TDTBsubI*. The SS, TT, and FF global process corners of *TDTBsub* were tested from V_{dd} 0.2 V to 1.2 V (without Monte-Carlo). The circuit passed all SS, TT, and FF corner variations with the regards to the testing plan of Section 5.1. A more robust approach to taking global variations into account is found by running a Monte-Carlo simulation at each global process corner [2]. This idea is applied for *TDTBsubII*.

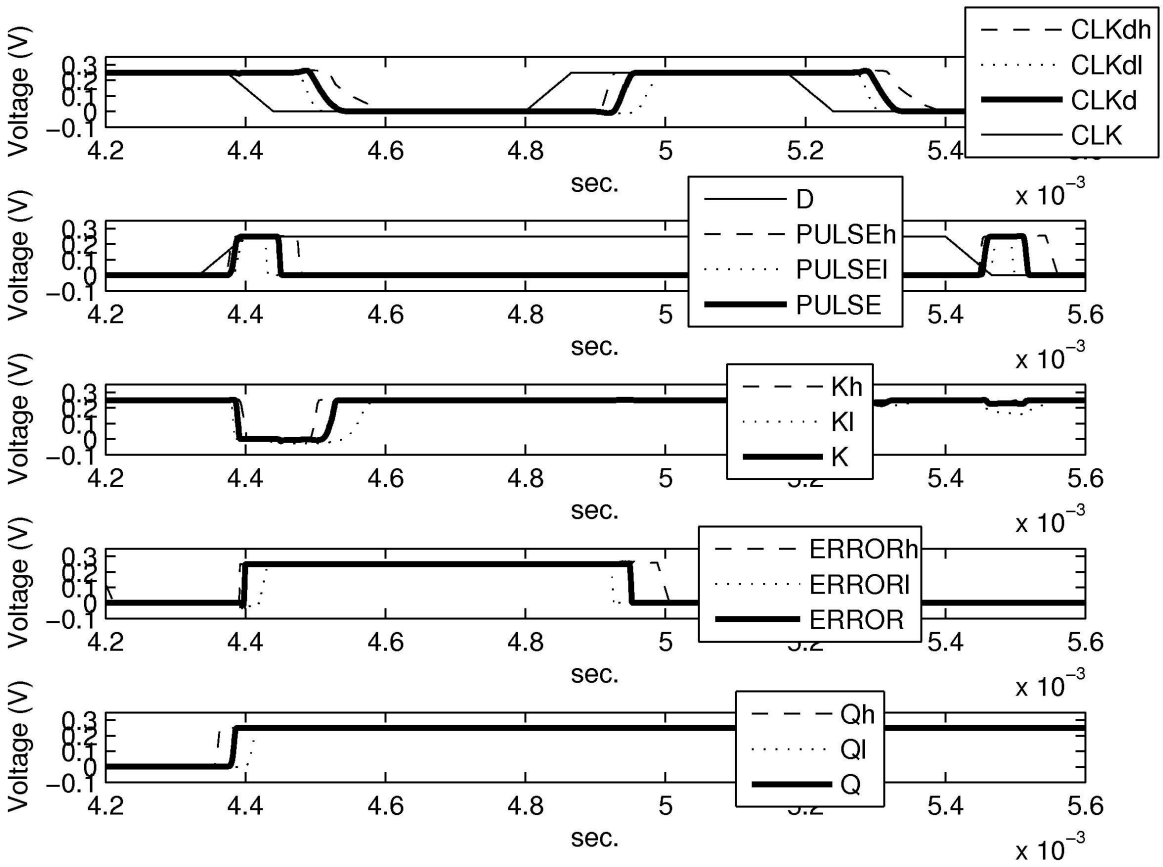


Figure 5.6: 1000 point Monte-Carlo Simulation at 0.25 V of *TDTBsubI*. When *D* arrives late, an *ERROR* signal results. Transitions that occur close to the *CLK* edge are sensitive to variations in *CLKd* and *PULSE*.

5.3 *TDTBsubII*

Simulations of *TDTBsubII* were performed to understand the operating frequency, energy per operation, and functionality with variations. The testing plan of Section 5.1 was applied to each simulation. The $t_{r,f}$ used for *D* and *CLK* within each simulation was calculated using the 5-stage ring oscillator in Section 4.3.4. *TDTBsubII* is capable of operation from 0.3 V to 1.2 V.

5.3.1 Operating Frequency

A simulation to understand the operating frequency of *TDTBsubII* was performed using the TT global process corner from 0.3 V to 1.2 V and a sampling resolution of 0.05 V. The results are shown in Fig. 5.7. Similarly to *TDTBsubI*, the results closely follow the curves previously predicted by the delay equations (3.5) and (3.6) in Section 3.1.1. A frequency of 120 MHz was found at V_{dd}=1.2 V while a minimum frequency of 2.78 kHz was found at V_{dd}=0.3 V. Below 0.3 V, *TDTBsubII* does not operate correctly due to increased leakage.

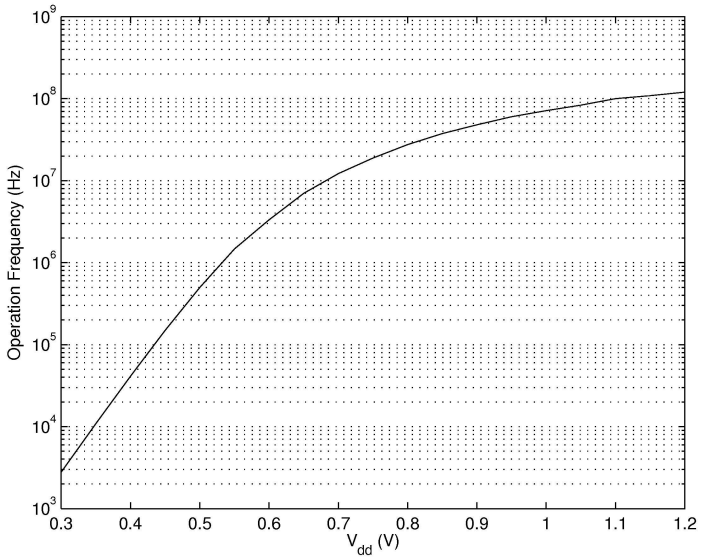


Figure 5.7: Operating frequency of *TDTBsubII* as V_{dd} is swept.

5.3.2 Energy Consumption

Fig. 5.8 shows the total average energy per operation (E_{TOTop}) as V_{dd} is swept from 0.3 V to 1.2 V for *TDTBsubII*. The MEP is near V_{dd}=0.35 V which is close to LATCH1's MEP of V_{dd}=0.4 V. This indicates that E_{LEAKop} is a similar percentage of E_{TOTop} for *TDTBsubII* and LATCH1. By comparing Fig. 5.8 and Fig. 5.4, it is can be calculated that *TDTBsubII* consumes about 8 and 14 times more E_{TOTop} than LATCH1 at 0.3 V and 1.2 V, respectively.

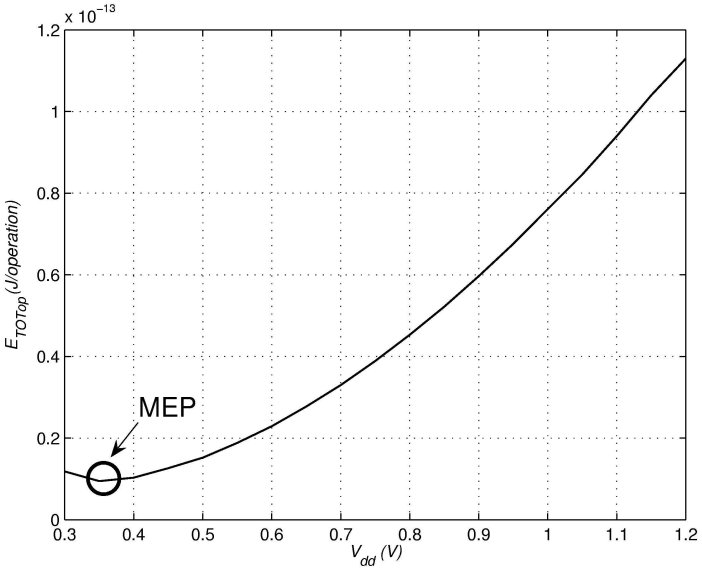


Figure 5.8: Total average energy consumption per operation (E_{TOTop}) for a *TDTBsubII* with a MEP at 0.35 V.

To further understand the amount of energy overhead used in adding TED capability to LATCH1, E_{TOTop}/E_{L1op} was constructed as shown in Fig. 5.9. As V_{dd} grows from 0 V to 1.2 V, the amount of energy overhead grows larger. The E_{TOTop} averaged over all V_{dd}

shows that *TDTBsubII* consumes about 13.8 times more energy than LATCH1. Although adding TED capability to LATCH1 increases the energy consumption, system-level simulations have shown that significant energy savings can be achieved by placing TED-latches at all critical paths in a pipeline [10].

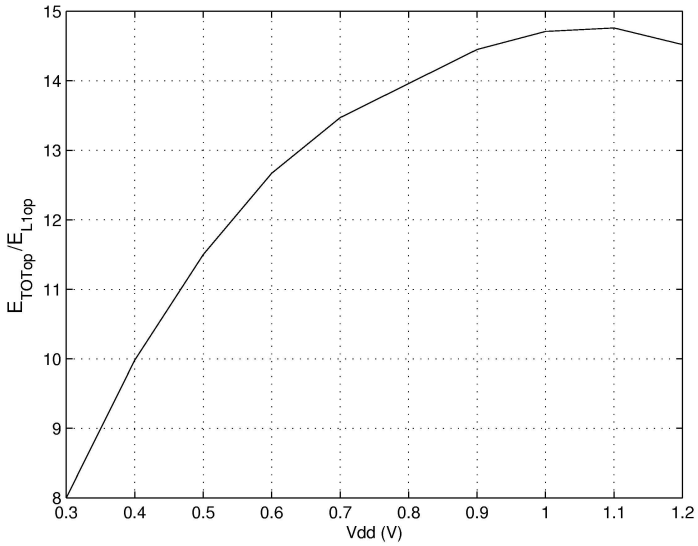


Figure 5.9: E_{TOTop}/E_{L1op} versus V_{dd}.

5.3.3 Functionality with Variations

Both local and global variations were taken into account for *TDTBsubII*. Local variations were first understood with a 1000 point Monte-Carlo simulation performed at the TT process corner from V_{dd} 0.2 V to 1.2 V. The results of the simulation at V_{dd}=0.35 V are displayed in Fig. 5.10. All three test cases for *D* transitions are from Section 5.1.

In addition to local variations, global variations were briefly examined by running a Monte-Carlo simulation at the following global process corners: TT, FF, SS. For each process corner, all three test cases for *D* transitions in Section 5.1 passed without any exceptions. This simulation provides a more robust approach to taking global variations into account while running Monte-Carlo simulations [2].

5.4 Comparison of *TDTBsubI* and *TDTBsubII*

Differences in operation frequency, energy consumption, and functionality with variations between *TDTBsubI* and *TDTBsubII* are reviewed in this section. A small difference in operation frequency was observed between the two *TDTBsub* versions. *TDTBsubI*'s minimum frequency of 2.5 kHz at 0.3 V is slightly less than *TDTBsubII* at 0.3 V (i.e. 2.78 kHz). The operation frequency at V_{dd}=1.2 V was 143 MHz and 120 MHz for *TDTBsubI* and *TDTBsubII*, respectively. These operation frequencies are slightly less than Razor II's operation frequency at 1.2 V (i.e. 185 MHz) [10]. However, the operation frequency of *TDTBsubI* and

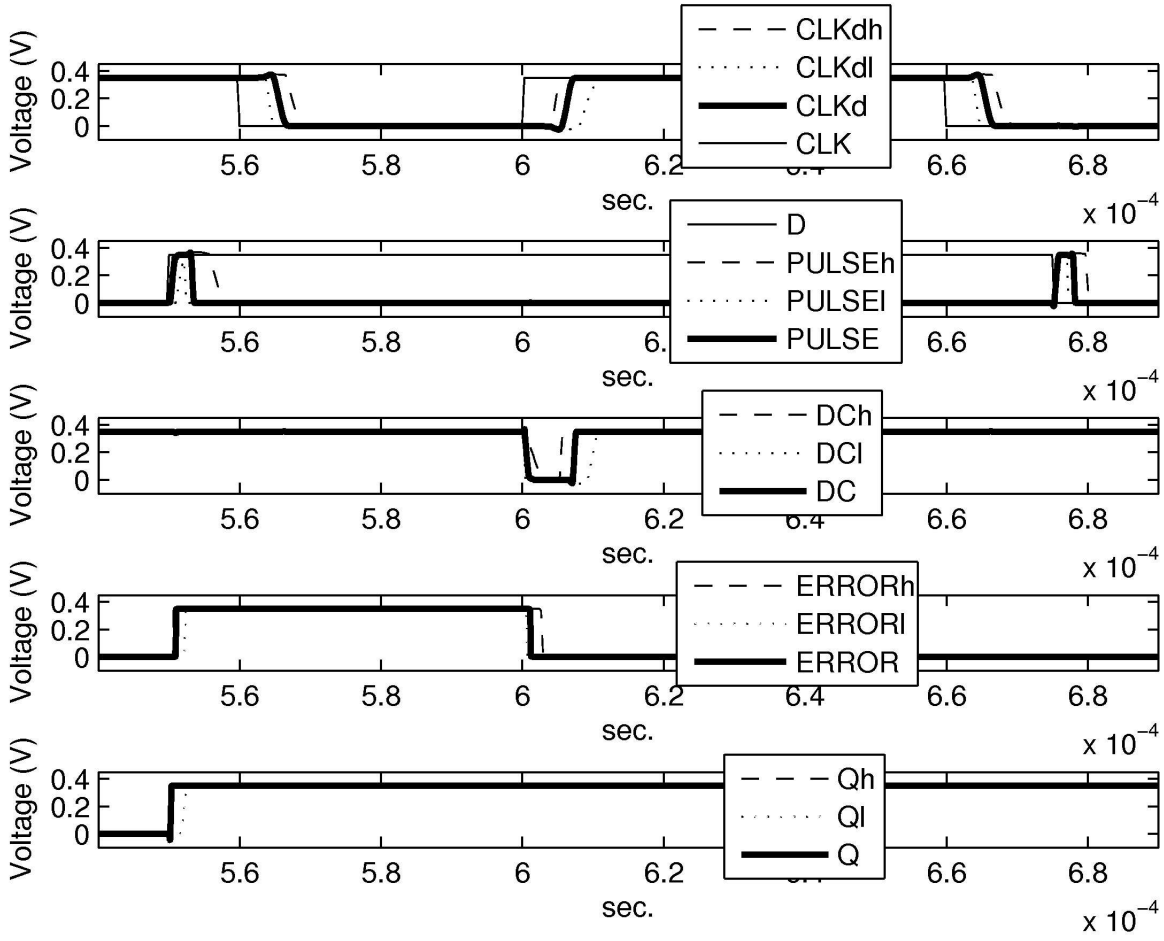


Figure 5.10: 1000 point Monte-Carlo Simulation at V_{dd}=0.35 V of *TDTBsubII*. When *D* arrives late, an *ERROR* signal results.

TDTBsubII could be increased by making the size of t_{dc} from the testing plan (Section 5.1) larger.

The comparison of energy consumption between the two *TDTBsubI* versions showed a large difference. As displayed in Fig. 5.11, *TDTBsubII* consumes less energy over the entire range of V_{dd}. The primary reason for a difference in energy consumption is the *CLK* delay. The emulation inverters (i.e. inverter 5, 6, and 7 from Fig. 4.3) were not used in *TDTBsubII*. Additionally, the inverters used to delay the *CLK* signal in *TDTBsubII* were sized smaller. The cost for reduced energy consumption in *TDTBsubII* is a reduction in operating range. Smaller device sizes for *TDTBsubII* resulted in larger variations during Monte-Carlo simulations. *TDTBsubII* was able to operate correctly to V_{dd}=0.3 V while *TDTBsubI* functioned to V_{dd}=0.2 V.

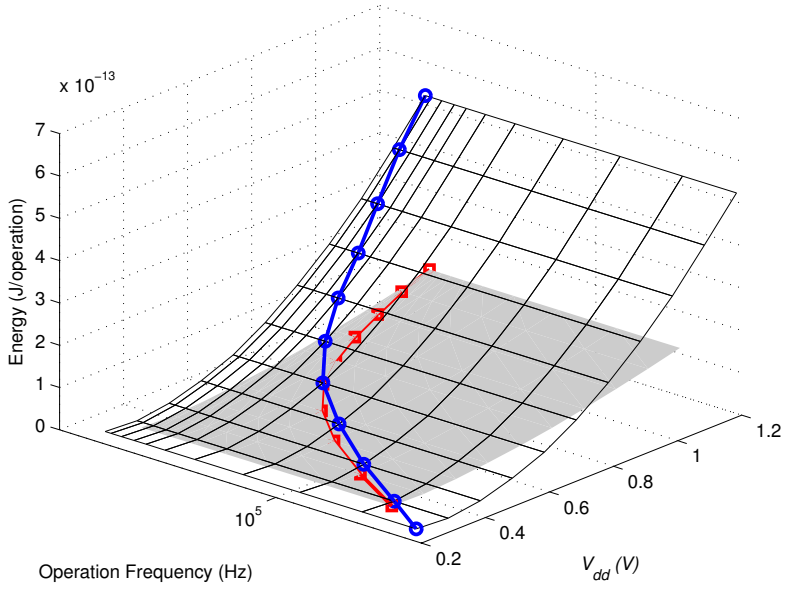


Figure 5.11: Energy consumption of both *TDTBsub* versions.

5.5 System-level Test Circuit

To understand the operation of *TDTBsub* at the system level, a modified version of *SystemTest1* (Section 4.4) was constructed (Fig. 5.12). It is called *SystemTest2* and it used to test the use of TED on an adder. The input a<0> to a<n-1> and b<0> to b<n-1> is first loaded into the adder. After adding the input (e.g. a<0>+b<0>, a<1>+b<1>, ... a<29>+b<29>), the output is passed to the *TDTBsubI* latches. *TDTBsubI* latches were used for the simulation since the submitted chip was built with this version. If the data becomes too slow through the adder due to the V_{dd} level, the *TDTBsubI* latches generate *ERROR* signals.

A new version of the adder from *SystemTest1* was constructed in *SystemTest2* to amplify the effects of delay on TED. The modified adder is a 30-bit ripple carry adder (Fig. 5.13). Each bit position is represented by a full-adder. The addition is performed from the least significant bit (LSB) to the most significant bit (MSB). A carry bit propagates from right to left. Each full-adder gives a delay δt before its outputs s<n> and c<n+1> are valid for the next stage. The smallest delay is found when there is no carry bits. The largest delay results when a carry bit propagates through each FA stage. For this case, the output at the MSP needs to wait 30*δt to have a valid output [34].

5.5.1 *SystemTest2* Functionality

To verify the functionality of *SystemTest2*, two input test cases were simulated (Fig. 5.14). The simulations were both performed at V_{dd}=0.3 V under typical process parameters. The data to be added was loaded at the rising edge of the *CLK* signal. Test case A. shows the

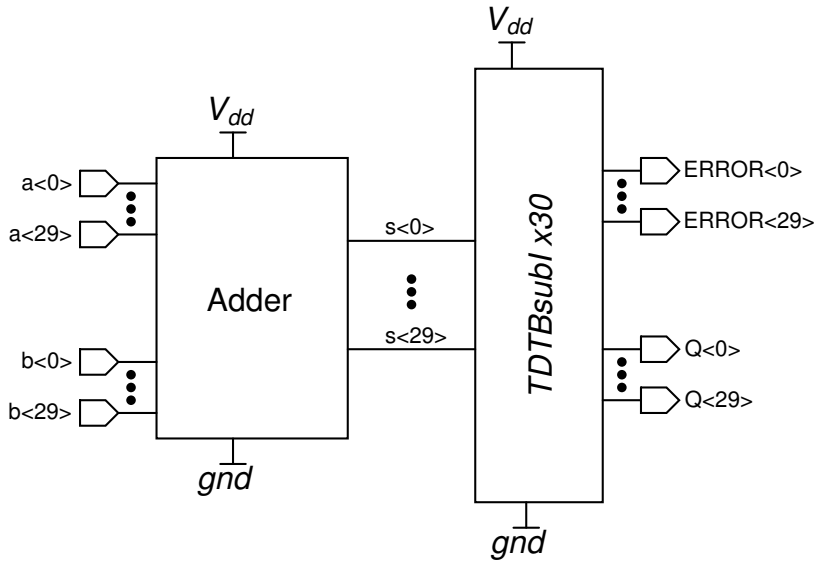


Figure 5.12: *SystemTest2*.

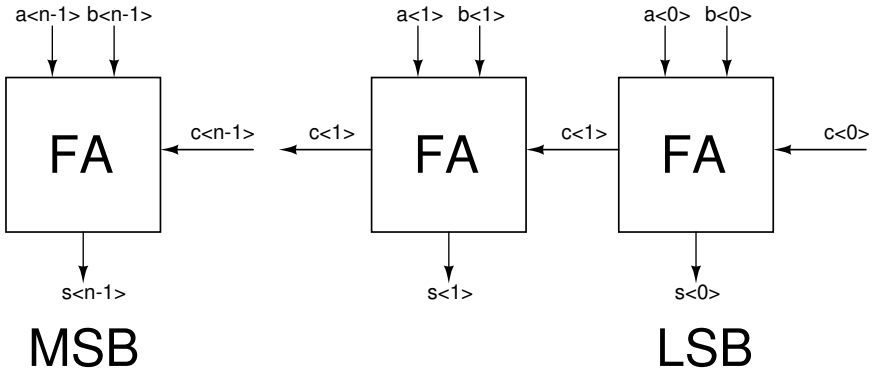
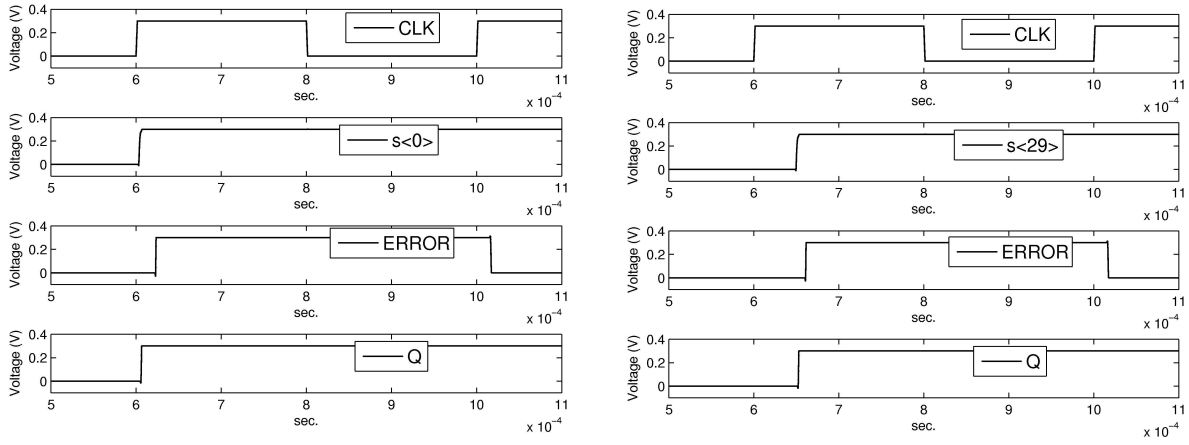


Figure 5.13: Ripple carry adder.

combination of input data that gives the minimum amount of carry bits and thus the smallest delay through the adder (i.e. [00 ... 01] + [00 ... 00]). Test case B. gives the longest delay possible due to the large amount of carry bits (i.e. [01 ... 11] + [00 ... 01]). An *ERROR* is generated for each test case since the output of the adder transitions under the *CLK* high of *TDTBsubI*.

To be considered valid during one TED sampling period, both s<0> and s<29> should be larger than T_{dmin} yet smaller than T_{dmax} (Fig. 5.15). To operate without *ERROR* signals, the adder output should fall within the T_{dmeet} region. This region ensures that V_{dd} is operating at a correct level to meet any processing tasks (i.e. addition). More specifically, the adder's output should be designed to fall at the rightmost point of the T_{dmeet} region to have minimum energy consumption. To attain transitions of s<0> and s<29> within T_{dmeet}, two conditions must be met. First, the smallest delayed output of the adder (i.e. T_{Amin}) should have a longer delay than T_{dmin}. Otherwise, incorrect *ERROR*s are generated from the previous T_{valid} region. Second, the adder's longest delayed output T_{Amax} is required to be less than



(a) (b)

Figure 5.14: *SystemTest2* simulations (a.) Test case A. (b.) Test case B.

T_{dmin}+T_{dmeet}. For the test cases in Fig. 5.14, T_{Amin} and T_{Amax} are equal to s<0> and s<29>, respectively. Variations at the edge of T_{Amin} and T_{Amax} must also be considered as explained in the next section.

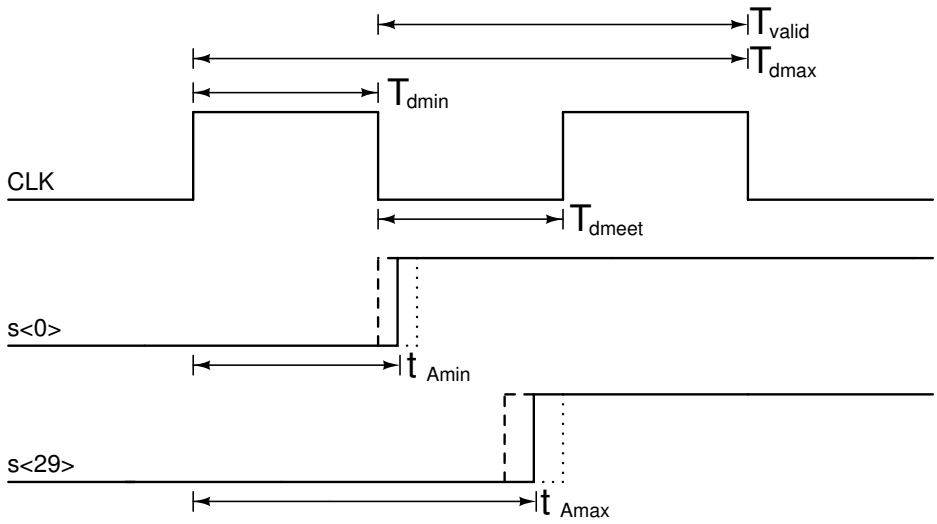


Figure 5.15: Minimum and maximum path delays where T_{dmin} is the minimum path delay, T_{dmax} is the maximum path delay, and T_{dmeet} is the delay range that guarantees error free operation. T_{Amin} is minimum delay of the adder while T_{Amax} is the maximum.

5.5.2 *SystemTest2* Functionality with Variations

As V_{dd} is reduced, variations have a large effect on the operation of *SystemTest2*. For valid operation of *SystemTest2*, both T_{Amin} and T_{Amax} must fit into the T_{valid} region even with variations that alter the delay of the adder. Since T_{Amin} uses the least amount of full-adders to compute the final sum, it has less variations in delay. The variations in delay of T_{Amax} are

larger since the final sum is calculated using the maximum number of full-adders. Thus, the variations in each full-adder add the to overall delay variation.

A 1000-point Monte-Carlo simulation was performed on *SystemTest2*. The width of T_{Amin} variations are approximately 10μs at V_{dd}=0.3 V. Assuming a Gaussian distribution, T_{Amin} should be delayed by 5μs past T_{dmin} to ensure that *ERROR*s are not incorrectly generated. For T_{Amax}, the variations at V_{dd}=0.3 V were approximately 80μs. As expected, the effect of variations was lager on T_{Amax}. To operate within T_{dmeet}, T_{Amax} should be designed to transition a minimum of 40μs before T_{dmin}+T_{dmeet}.

To understand the effects on variations as V_{dd} is reduced, another Monte-Carlo simulation was performed on *SystemTest2* at both V_{dd}=1.2 V and 0.3 V. Fig. 5.16 shows the *ERROR* rate percentage as V_{dd} is reduced by steps of 0.005 V. For each V_{dd} step reduction, a 1000-point Monte-Carlo simulation was performed. As V_{dd} is reduced from 1.2 V, the ERROR rate is negligible until about 1.185 V. Next, the same simulation was performed at V_{dd}=0.3 V with the same size steps in V_{dd} reduction as for 1.2 V (Fig. 5.17). The local variations are much larger for sub-threshold and thus cause a higher *ERROR* rate as V_{dd} is reduced.

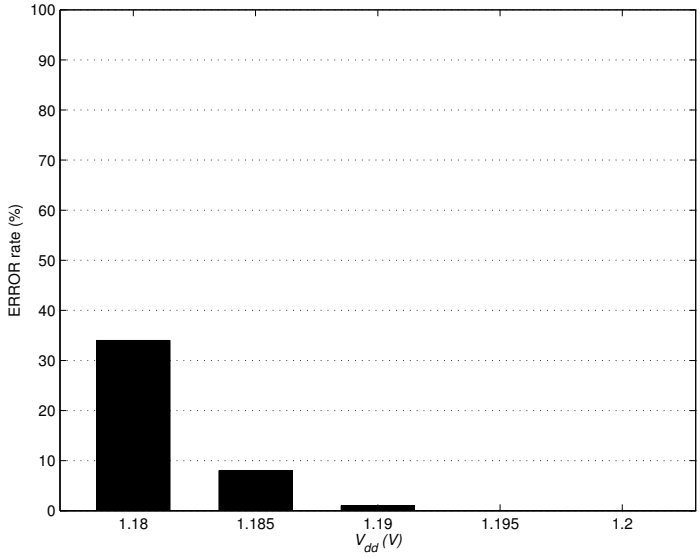


Figure 5.16: ERROR rate (%) as V_{dd} is lowered from 1.2 V.

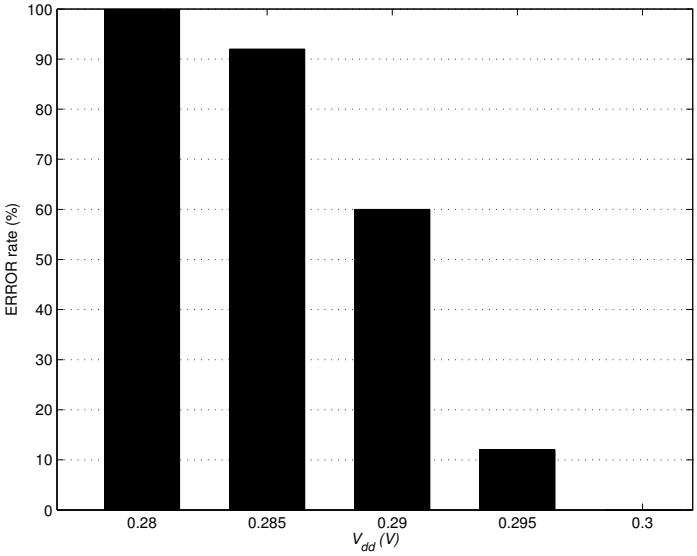


Figure 5.17: ERROR rate (%) as V_{dd} is lowered from 0.3 V.

Chapter 6

Conclusions and Future Work

This thesis first explained the concept of a dynamic voltage scaling (DVS) system since timing error detection (TED) is used within such a system. Additional detail about the operation of TED and previous TED implementations were then given. Next, the theory of static CMOS was given to highlight the challenges faced within the sub-threshold region. The design of both *TDTBsub* latch versions and a system-level circuit that used *TDTBsub* latches was then presented. As expected from theory, most of the time needed to design the TED latches was in sub-threshold. A section giving the simulation results of all designed *TDTBsub* components was then provided.

Simulations showed that the goal of the thesis (i.e. to operate TDTB in sub-threshold) was attained. The operation voltage of both *TDTBsub* versions worked deep into sub-threshold. The operation frequency slowed significantly as V_{dd} was scaled to sub-threshold with the benefit of operation at the minimum energy point (MEP). Monte-Carlo simulations were successful in proving the robust operation of *TDTBsub* in sub-threshold. This was the result of increased transistor sizes and logic style choices. Fig. 5.11 highlighted the fact that although *TDTBsubI* had excessive energy consumption in strong inversion, its energy consumption in sub-threshold was comparable to *TDTBsubII*. In *TDTBsubII*, the *CLK* delay is the same but is generated using much smaller inverters. Considering that the pull-down network and keeper circuit operate identical and are sized similarly in both *TDTBsub* versions, it can be concluded that the measurement results from *TDTBsubI* are valuable in sub-threshold. Additionally, the new idea in *TDTBsubII* used to reset the *ERROR* signal low, proved to be an effective method.

The layout of *TDTBsubI* and a system-level test circuit were constructed in 65 nm CMOS. The *TDTBsubII* latch was not built since it was designed after the chip deadline. A PCB board was also built to test the chip. Upon inspection, the chip was determined to be inoperative. The final metal layer (metal 6) was not placed on the chip during processing. This prevented a connection from the pad ring circuit to the *TDTBsubI* latch and adder circuit. This mistake was a result of the manufacturering process and not the design in this work.

Although the goal of this thesis was attained, additional work is recommended. First,

it is worthwhile to identify and simulate more global process corners at each Monte-Carlo simulations. *TDTBsubII* was simulated with a number of different global process corners but more could be added. System-level simulations have shown that significant energy savings can be achieved by placing TED-latches at all critical paths in a pipeline [10][11]. However, this concept needs to be proved in sub-threshold. Using the *TDTBsub* latch in a real processor could provide more accurate results. To improve the energy consumption, it is recommended to reduce the size or eliminate the *CLK* delay inverters. This could be addressed through system-level constraints or a new TED latch. A new TED latch without the requirement for such an accurate *PULSE* signal would be the best option. Finally, an effective method to tune the supply voltage (V_{dd}) with TED is recommended.

Bibliography

[1] G. J. Pendock, L. Evans, and G. Coulson, "Wireless sensor module for habitat monitoring," in *Proc. 3rd International Conference on Intelligent Sensors, Sensor Networks and Information ISSNIP 2007*, 2007, pp. 699–702.

[2] A. C. A. Wang, B. Calhoun, *Sub-Threshold Design for Ultra Low-Power Systems*. Springer, 2005.

[3] S. Henzler, *Power Management of Digital Circuits in Deep Sub-Micron CMOS Technologies*. Springer, 2007.

[4] M. Elgebaly and M. Sachdev, "Variation-aware adaptive voltage scaling system," *IEEE Trans. VLSI Syst.*, vol. 15, no. 5, pp. 560–571, 2007.

[5] S. Akui, K. Seno, M. Nakai, T. Meguro, T. Seki, T. Kondo, A. Hashiguchi, H. Kawahara, K. Kumano, and M. Shimura, M. A10 Shimura, "Dynamic voltage and frequency management for a low-power embedded microprocessor," in *Proc. Digest of Technical Papers Solid-State Circuits Conference ISSCC. 2004 IEEE International*, K. Seno, Ed., 2004, pp. 64–513 Vol.1.

[6] Y. Ramadass and A. Chandrakasan, "Minimum energy tracking loop with embedded dc-dc converter delivering voltages down to 250mv in 65nm cmos," in *Proc. Digest of Technical Papers. IEEE International Solid-State Circuits Conference ISSCC 2007*, A. Chandrakasan, Ed., 2008, pp. 64–587.

[7] M. Najibi, M. Salehi, A. Kusha, M. Pedram, S. Fakhraie, and H. Pedram, "Dynamic voltage and frequency management based on variable update intervals for frequency setting," in *Proc. IEEE/ACM International Conference on Computer-Aided Design ICCAD '06*, M. Salehi, Ed., 2006, pp. 755–760.

[8] J. Tschanz, N. S. Kim, S. Dighe, J. Howard, G. Ruhl, S. Vanga, S. Narendra, Y. Hoskote, H. Wilson, C. Lam, C. A10 Lam, M. Shuman, M. A11 Shuman, C. Tokunaga, C. A12 Tokunaga, D. Somasekhar, D. A13 Somasekhar, S. Tang, S. A14 Tang, D. Finan, D. A15 Finan, T. Karnik, T. A16 Karnik, N. Borkar, N. A17 Borkar, N. Kurd, N. A18 Kurd, and V. De, V. A19 De, "Adaptive frequency and biasing techniques for tolerance to dynamic temperature-voltage variations and aging," in *Proc. Digest*

of Technical Papers. IEEE International Solid-State Circuits Conference ISSCC 2007, N. S. Kim, Ed., 2007, pp. 292–604.

[9] K. A. Bowman, J. W. Tschanz, N. S. Kim, J. C. Lee, C. B. Wilkerson, S.-L. L. Lu, T. Karnik, and V. K. De, "Energy-efficient and metastability-immune timing-error detection and instruction-replay-based recovery circuits for dynamic-variation tolerance," in *Proc. Digest of Technical Papers. IEEE International Solid-State Circuits Conference ISSCC 2008*, 2008, pp. 402–623.

[10] S. Das, C. Tokunaga, S. Pant, W. H. Ma, S. Kalaiselvan, K. Lai, D. M. Bull, and D. T. Blaauw, "Razorii: In situ error detection and correction for pvt and ser tolerance," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 32–48, 2009.

[11] K. A. Bowman, J. W. Tschanz, N. S. Kim, J. C. Lee, C. B. Wilkerson, S.-L. L. Lu, T. Karnik, and V. K. De, "Energy-efficient and metastability-immune timing-error detection and instruction-replay-based recovery circuits for dynamic-variation tolerance," in *Proc. Digest of Technical Papers. IEEE International Solid-State Circuits Conference ISSCC 2008*, 2008, pp. 402–623.

[12] T. Austin, D. Blaauw, T. Mudge, and K. Flautner, "Making typical silicon matter with razor," *IEEE Computer Society*, vol. 37, no. 3, pp. 57–65, 2004.

[13] S. E. Wang, Alice; Naffziger, *Adaptive Techniques for Dynamic Processor Optimization*. Springer, 2008.

[14] L. Anghel and M. Nicolaidis, "Cost reduction and evaluation of a temporary faults detecting technique," in *Proc. Design Automation and Test in Europe Conference and Exhibition 2000*, 2000, pp. 591–598.

[15] D. Blaauw, "Razor ii: In situ error detection and correction for pvt and ser tolerance," *ISSCC 2008*, 2008.

[16] D. Ernst, N. S. Kim, S. Das, S. Pant, R. Rao, T. Pham, C. Ziesler, D. Blaauw, T. Austin, K. Flautner, and T. Mudge, "Razor: a low-power pipeline based on circuit-level timing speculation," in *Proc. 36th Annual IEEE/ACM International Symposium on MICRO-36 Microarchitecture*, 2003, pp. 7–18.

[17] C. A. Rabaey, J., *Digital Integrated Circuits*. Prentice Hall, 2003.

[18] J. R. Baker, *CMOS: Circuit Design, Layout, and Simulation*. IEEE Press, 2007.

[19] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer cmos circuits," *Proc. IEEE*, vol. 91, no. 2, pp. 305–327, 2003.

[20] A. Tajalli, E. J. Brauer, Y. Leblebici, and E. Vittoz, "Subthreshold source-coupled logic circuits for ultra-low-power applications," *IEEE J. Solid-State Circuits*, vol. 43, no. 7, pp. 1699–1710, 2008.

[21] E. C., "A short story of the ekv mos transistor model," *SSCS: IEEE Solid-state circuit society news*, vol. 13, pp. pp. 24–29, 2008.

[22] H. D. Allen, P., *CMOS Analog Circuit Design*. Oxford, 2002.

[23] C. A. Narendra, Siva G., *Leakage in Nanometer CMOS Technologies*. Springer, 2006.

[24] N. Sirisantana and K. Roy, "Low-power design using multiple channel lengths and oxide thicknesses," *IEEE Design & Test of Computers*, vol. 21, no. 1, pp. 56–63, 2004.

[25] T. Karnik, "Variation-tolerant circuit design techniques. slide: Sources of variability." in *ISSCC Forum 6*, 2008.

[26] K. A. Bowman, S. G. Duvall, and J. D. Meindl, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *IEEE J. Solid-State Circuits*, vol. 37, no. 2, pp. 183–190, Feb. 2002.

[27] A. Bellaouar, A. Fridi, M. J. Elmasry, and K. Itoh, "Supply voltage scaling for temperature insensitive cmos circuit operation," *IEEE Trans. Circuits Syst. II*, vol. 45, no. 3, pp. 415–417, 1998.

[28] J. Kwong, Y. K. Ramadass, N. Verma, and A. Chandrakasan, "A 65 nm sub-threshold microcontroller with integrated sram and switched capacitor dc-dc converter," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 115–126, 2009.

[29] C. Hwang and Y.-C. Cheng, "Use of lambert w function to stability analysis of time-delay systems," in *Proc. American Control Conference the 2005*, 2005, pp. 4283–4288 vol. 6.

[30] S. Rusu, S. Tam, H. Muljono, D. Ayers, and J. Chang, "A dual-core multi-threaded xeon processor with 16mb l3 cache," in *Proc. Digest of Technical Papers. IEEE International Solid-State Circuits Conference ISSCC 2006*, 2006, pp. 315–324.

[31] M. Nomura, Y. Ikenaga, K. Takeda, Y. Nakazawa, Y. Aimoto, and Y. Hagihara, "Delay and power monitoring schemes for minimizing power consumption by means of supply and threshold voltage control in active and standby modes," *IEEE J. Solid-State Circuits*, vol. 41, no. 4, pp. 805–814, 2006.

[32] A. Chavan and E. MacDonald, "Ultra low voltage level shifters to interface sub and super threshold reconfigurable logic cells," in *Proc. IEEE Aerospace Conference*, 2008, pp. 1–6.

[33] M. Young. (2009, March) Bsim homepage. BSIM Research Group.

[34] V. Z. Brown, S., *Fundementals of Digital Logic with VHDL Design*. McGraw Hill, 2000.

Appendix A

TDTBsubI layout

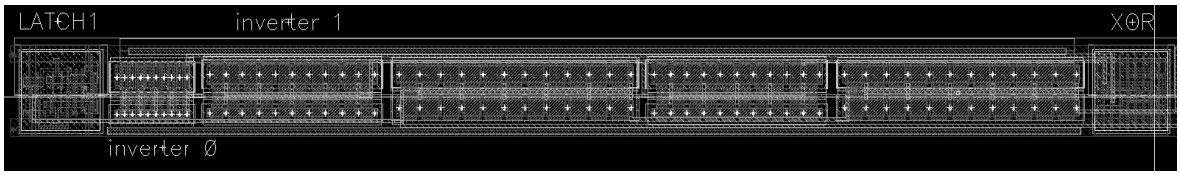


Figure A.1: Left half of *TDTBsubI* layout.

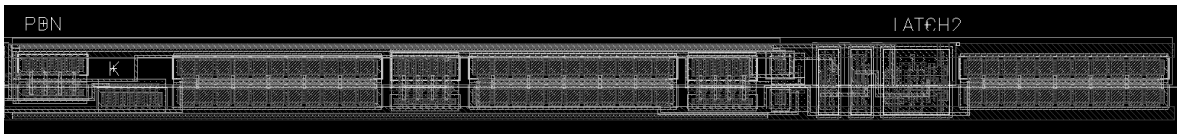


Figure A.2: Right half of *TDTBsubI* layout.

Appendix B

Level-shifter layout

Figure B.1: Level-shifter layout.

Appendix C

Photomicrograph of entire chip

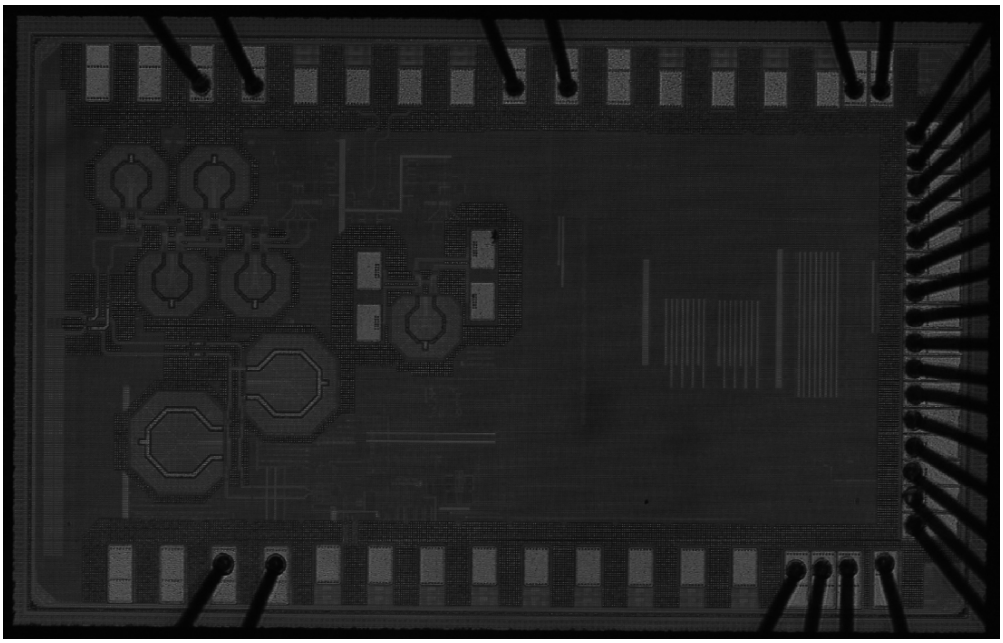


Figure C.1: Photomicrograph of entire chip.